

Money counts for a Times Higher Education top-rank

Citation for published version (APA):

Marconi, G., & Ritzen, J. (2014). *Money counts for a Times Higher Education top-rank*. Maastricht University, Graduate School of Business and Economics. GSBE Research Memoranda No. 015 <https://doi.org/10.26481/umagsb.2014015>

Document status and date:

Published: 01/01/2014

DOI:

[10.26481/umagsb.2014015](https://doi.org/10.26481/umagsb.2014015)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Gabriele Marconi, Jo Ritzen

**Money counts for a Times
Higher Education top-rank**

RM/14/015

GSBE

Maastricht University School of Business and Economics
Graduate School of Business and Economics

P.O. Box 616
NL- 6200 MD Maastricht
The Netherlands

Money counts for a Times Higher Education top-rank

– G. Marconi*, J. M. M. Ritzen** – April 2014

* School of Business and Economics, Maastricht University (NL), g.marconi@maastrichtuniversity.nl

** UNU-MERIT, Maastricht (NL), jo.ritzen@empowereu.org

Abstract

This paper analyses the relationship between a university's expenditure per student and its position in international university rankings. We take into account other factors that are expected to play a role, such as university mission, size, and productive inefficiency. We formalise these concepts in our theoretical model of rankings and universities, and estimate this model with data on universities classified in the top 200 by the Times Higher Education Supplement ranking of 2007. We find that the elasticity of a university's ranking score for the expenditure per student is equal to 8.9%, and that there are no clear signs of inefficiency in production among these universities. University mission and size are also significant predictors of ranking score. These results are important in view of the relevance attributed to rankings by government officials, university directors and students.

Keywords: expenditure – universities – international rankings – THES – production function

JEL Classification: C51 – D20 – F23 – I23

Acknowledgments:

We thank Andries de Grip, Frank Cörvers, Carla Haelermans, Caren Butter, Martijn Jungst, and the participants at the O&S Lunch Seminar in Maastricht (NL, 24/10/2012), at the AiO Seminar in Maastricht (NL, 20/08/2013), and at the ISNE 2013 conference in Maynooth (IE, 12/09/2013) for helpful comments and suggestions.

1. Introduction

International university rankings have captivated the attention of many decision-makers in higher education. Students look at these rankings when choosing which university to attend, employers take them into consideration when recruiting employees, and other stakeholders use them for decisions about funding and sponsorships (Hazelkorn, 2008). The reason is probably that, despite criticisms on their methodology (e.g. Marginson, 2007; Saisana, d’Hombres, and Saltelli, 2011), rankings are the only broadly available comparative measure of higher education institutional performance at an international level.

University rankings are considered a measure of education “quality”, which is on the agenda of policy-makers as it is considered to contribute to the aggregate level of human capital (Hanushek and Woessmann, 2012). As a result, rankings can be used for benchmarking the performance of universities and higher education systems. Another reason why government officials and university directors are interested in university rankings is that they have been shown to influence the preferences of students, not only at a national level (Dichev, 1999; Monks and Ehrenberg, 1999; Sauder and Lancaster, 2006) but also at an international level (Marconi, 2013). The market for international students has been growing dramatically in the last few decades, with the number of students enrolled outside their country of citizenship rising from 0.6 million worldwide in 1975 to 2.9 million in 2006 (OECD, 2008). Governments are interested in gaining shares of this market, because international students often end up working (at least temporarily) in their host country after completing their studies (Bijwaard, 2010). University leadership are interested in international students sometimes because of tuition fees , sometimes because they can increase the overall quality of the student body in a university and sometimes because they contribute to cross cultural learning .

Governments and university leadership generally want to improve the standing of a university in international university rankings. The most intuitive way of improving this standing is by increasing the resources available to a university. This should translate into better (or more) teachers, researchers, and facilities, which should improve the indicators of quality on which rankings are based. However, it is possible that the university is inefficient in using its resources, or that the university prefers to spend them on dimensions of performance that are not considered important by the ranking institutions. In that case, the effect of expenditure on the standing in university rankings would be hampered.

In this paper, we argue that a university's ranking score is closely related to per student expenditure. We also take into account other factors that play a role, such as (in)efficiency and the mission of a university. We formalise these considerations in our model, where universities are multi-output organisations that maximise a combination of outputs based on their own utility function (the "mission"). The institutions are then assessed by a ranker, who assigns them a score based on subscores. We would observe perfect correlation between expenditures and the score if three conditions are fulfilled: 1) the ranker measures the outputs perfectly in the subscores; 2) universities are efficient; 3) universities pursue the same mission, so that the weights given to the subscores reflect the identical missions of the universities. If some of these conditions do not hold, we show that per student expenditure affects the rank positively conditional on the university's mission, while inefficiency affects a university's score negatively.

We regress this composite score on expenditure and other control variables, estimating a single-product production function in which the dependent variable is a function of different outputs, and the only input is expenditure. This approach, while suitable for addressing our research question, differentiates our paper from other work studying the input-output relationships in higher education. In the literature in the field we usually find estimated cost functions with several performance indicators, which are not aggregated (Longlong et al., 2009; Lenton, 2008; Kuo Ho, 2008; Johnes and Johnes, 2009; Izadi et al., 2002).

We test our model with data on university expenditure, the number of students and other institutional characteristics for 155 of the 201 universities that have been ranked as the top 200 in the world by the Times Higher Education Supplement – Quacquarelli-Sismonds (THES-QS) ranking in 2007. The sample size is extended to 183 universities when we exclude some control variables. The data has been collected for an earlier research project (see Ritzen, 2010, sec. 2.2), which constitutes the starting point for this enquiry. The rankings used in this paper are described in Ince (2007) and O'Leary et al. (2008).

Our findings show a good model fit, and our estimate of the elasticity of the score for expenditure per student is 9%. This implies that a university gains one position in the rank by increasing its expenditure per student by 3%. Our proxy variables for university mission are also significant predictors of university score. Using a test suggested by Coelli (1995), we find no evidence of productive inefficiency in our sample. Our findings are robust to taking unobserved heterogeneity into account (by including geographical dummies and the lagged dependent variable).

The paper is structured as follows. Section 2 discusses the basic problem of the evaluation of the output of a firm when competitive prices do not exist or are not observed, and describes our model. Section 3 describes the data. Section 4 shows the results of the empirical estimations. Section 5 concludes.

2. Model

Universities can be seen as multi-output enterprises. Assessing the value of the outputs of a multi-outputs firm is, in principle, not a problem. The multiple outputs are aggregated in one index, the “value of the production”, by multiplying each output with the price of that output. As Fried et al. (2008, p. 9) point out, “market prices produce a natural set of weights” to aggregate the outputs. The value added is then the value of the production at market prices minus the costs incurred in the purchase of the inputs. It is then possible to compute measures of productivity by dividing this value by the amount of inputs used.

When prices are not observed or do not reflect competitive outcomes, aggregation of multiple outputs into a composite indicator becomes problematic. A set of weights must be chosen instead of the prices. In the last decades the proliferation of composite indicators of university production based on different sets of weights has been the result. The approach for aggregating multiple outputs into one followed by the THES is such that the weights are assigned without a specific relation to the observed data, but are based on the opinion or on the preferences of an external observer. This approach follows from the fact that ranking institutions are interested in organisational performance, but not necessarily in performance conditional on organisational preferences. As a result, it is the ranking institution which chooses the weights, which are not endogenously determined by the outputs or the inputs of the production process. This has a theoretical justification: inputs and outputs depend on the objective function of the university, which is not necessarily relevant to other stakeholders.

Even when the weights are not chosen based on the university’s preferences, the ranking should be a positive function of the expenditure the university incurred purchasing the inputs as long as both the ranker’s score and the university utility function are monotonic functions of the outputs. Since the ranker’s score and the university utility function are not the same, other factors will play a role as well, for example the mission of the university. We show this by using a Cobb-Douglas function for production, utility and ranking score.

Every university has a different utility function and maximises its utility, W_i . The utility is a function of the n outputs that can be produced by the university:

$$(1) W_i = w_i(O_{i1}, \dots, O_{in}) = O_{i1}^{\gamma_{i1}} \dots O_{in}^{\gamma_{in}}$$

Where O_{ij} is the amount of output of type j produced by university i , $j=1, \dots, n$; and $\gamma_i = [\gamma_{i1}, \dots, \gamma_{in}]$ is a vector of university-specific parameters determining the relative weight that is given to every quantity of output O_{i1}, \dots, O_{in} in the utility function. Hence, the vector γ_i determines the preferences of the university over different outputs that can be produced or, in other words, γ_i determines the mission of the university. The utility W_i could be taken to be the equivalent of the concept of prestige, which has been used in some studies as the appropriate objective function of universities (e.g. Cyrenne and Grant, 2009). It could, however, also reflect other objectives of the university, for example a genuine interest in serving under-represented groups. Notice that the utility function w_i is university-specific. This is different from the production functions that will be introduced, which are the same across universities.

The university produces n outputs by using m inputs. We assume that the production function is the same for all universities:

$$(2) \begin{cases} O_{i1} = f_1(x_{i1}, \dots, x_{im}, u_{i1}) = A_1 x_{i1}^{\alpha_{11}} \dots x_{im}^{\alpha_{1m}} e^{-u_{i1}} \\ \dots \\ O_{in} = f_n(x_{i1}, \dots, x_{im}, u_{in}) = A_n x_{i1}^{\alpha_{n1}} \dots x_{im}^{\alpha_{nm}} e^{-u_{in}} \end{cases}$$

Where x_{ik} represents the amount of input used k , $k=1, \dots, m$; $A = [A_1, \dots, A_n]$ and $\alpha = [\alpha_{11}, \dots, \alpha_{nm}]$ are parameters of the production function. Examples of outputs of a university are the number of graduates, or the articles published. This can be differentiated by discipline (e.g. graduates in the natural sciences, or number of publications in economics). Examples of inputs are the number of professors or computers. The production function can easily accommodate both economies of scale and economies of scope, since the f_j s are not required to be homogeneous of degree one and the same input can appear in the production process of an arbitrary number of outputs. u_{ij} is the term indicating how much the production of output j is decreased because of technical inefficiency of university i , $u_{ij} \geq 0$. Thus, the component u_{ij} represents possible inefficiency of the university in the production process, in line with the literature following Aigner et al. (1977). It represents the distance between the production of the institution and the output produced when efficient.

The ranker observes the university outputs imperfectly, and computes a score, S_i , based on the observed output measures \tilde{O}_{ij} . The observed output measure is assumed to be given by the multiplication of the real output O_{ij} by the exponential function of a stochastic measurement error term v_{ij} , $\tilde{O}_{ij} = O_{ij} \exp(v_{ij})$. The relative weights of the outputs in composing the ranking are contained

in the vector $\phi=[\phi_1,\dots,\phi_n]$. The ranker may be interested in the level of output relative to the size of the university, D_i . In this case, every output should be corrected for size. For example, the ranker may look at the human capital *per student*, or publications *per faculty*. For this reason we include the parameters θ_1,\dots,θ_n , reflecting the importance of the university size in the evaluation of each output in Equation (3).¹

$$(3) S_i = s(O_{i1}, \dots, O_{in}, v_{i1}, \dots, v_{in}, D_i) = \prod_{j=1}^n \frac{\delta_{ij}^{\phi_j}}{D_i^{\theta_j}} = \frac{\prod_{j=1}^n o_{ij}^{\phi_j} e^{v_{ij}}}{D^{\theta_T}}$$

Where $\theta_T = \sum_{j=1}^n \theta_j$.

The aggregation function in Equation (3) is similar to the one suggested by Fernández et al. (2000, 2002). In their Bayesian estimates of a multi-output production function in the banking and dairy sector, the authors aggregate the measured outputs of the firm through a Constant Elasticity of Substitution (CES) aggregation function. However, for analytical convenience, we choose a Cobb-Douglas aggregation function in this paper.

Substituting (2) into (1), we see that the total utility, W_i , is a Cobb-Douglas function of the inputs to production x_1,\dots,x_m . The university maximises Equation (1) subject to the budget constraint:

$$(4) B_i \geq \sum_{k=1}^m p_k x_{ik}$$

Where B_i is the total budget of university i and it is given exogenously (for an example of an endogenous budget, see Appendix A), and p_k is the unitary price of the input x_k . Taking the first-order maximisation conditions and manipulating these, we obtain the well-known result that expenditure on each input are a constant fraction of the total costs (e.g. Jehle & Reny, 2001, p. 130). This, in turn, is equivalent to say that the purchased quantity of an input, x_{ik} is proportional to the total budget of a university:

$$(5) x_{ik} = \frac{B_i \sum_{j=1}^n \alpha_{jk} \gamma_{ij}}{p_k \sum_{k=1}^m \sum_{j=1}^n \alpha_{jk} \gamma_{ij}}$$

Substituting Equation (5) into system of Equations (2) and then into (3), taking logs and manipulating yields:

$$(6) \ln[S_i] = \beta_0 + \beta_1 \ln \left[\frac{B_i}{D_i} \right] + \beta_2 \ln[D_i] + T_{1i} + T_{2i} + (v_i - u_i)$$

¹ We include size in the ranker's score function, but not in the university utility function. In reality, the university may care about relative measures of output as well. If the university size is exogenous, this is not a problem for our model, because size is not relevant in the university maximisation process. A case in which size is related to the budget of a university, so that size is endogenous, is considered in Appendix A. In this paper, size is proxied by the number of students.

Where:

$$\beta_0 \equiv \sum_{j=1}^n \phi_j \ln [A_j]$$

$$\beta_1 \equiv \left(\sum_{j=1}^n \phi_j \sum_{k=1}^m \alpha_{jk} \right)$$

$$\beta_2 \equiv \left[\left(\sum_{j=1}^n \phi_j \sum_{k=1}^m \alpha_{jk} \right) - \theta_T \right]$$

$$T_{1i} \equiv \frac{\sum_{k=1}^m \left(\sum_{j=1}^n \phi_j \alpha_{jk} \right) \cdot \ln [\sum_{j=1}^n \alpha_{jk} \gamma_{ij}]}{\left(\sum_{k=1}^m \sum_{j=1}^n \phi_j \alpha_{jk} \right) \cdot \ln [\sum_{k=1}^m \sum_{j=1}^n \alpha_{jk} \gamma_{ij}]}$$

$$T_{2i} \equiv - \sum_{k=1}^m \left(\sum_{j=1}^n \alpha_{jk} \phi_j \right) \ln [p_k]$$

$$(v_i - u_i) \equiv \sum_{j=1}^n \phi_j (v_{ij} - u_{ij})$$

Equation (6) is a linear equation for determining the log score that the ranker assigns to university i . The first term, β_0 , equals the sum over j of the logs of the parameters A_j , multiplied by the weight that the respective output receives in the computation of the overall score.

The second term is the log of expenditure per student, B_i/D_i , multiplied by the coefficient β_1 . This coefficient is the weighted sum of the parameters α_{jk} of the production functions, where every parameter of the production function of the j th output is multiplied by the weight that the j th output receives in the computation of the final score, ϕ_j . Interestingly, the coefficient remains positive as long as α_{jk} and ϕ_j are positive for every j and k . So, under the assumptions made for the functional specification, a positive value for this coefficient implies that investing more money increases the aggregated output of a university. The estimated coefficient, which is the same across universities, gives the elasticity of the ranker's score for the expenditure per student.

The third term is given by the log of the number of students, D_i , multiplied by β_2 . This coefficient is equal to the difference between the weighted sum of the α_{jk} parameters and the parameter θ_T .² The interpretation of β_2 is not straightforward. If it is positive, this could mean that there are increasing returns to scale to the production of at least some university outputs, or it could simply mean that for the ranker total outputs are more important than per-student outputs. Both the presence of economies of scale in the higher education production function and the possible existence of a "size bias" in international university rankings have been investigated in the literature (see Izadi et al., 2002; G. Johnes and Johnes, 2009; Longlong et al., 2009 for economies of scale in higher education;

² The value of θ_T would be equal to zero if the ranker does not correct output for the size of the institution, and it would be equal to the sum of the parameters ϕ_1, \dots, ϕ_n if the ranker is only interested in relative measures of output (i.e., output divided by the size of the university). In our baseline specification the estimate of θ_T is 0.015 (equal to the difference between β_1 and β_2). However, we do not know the parameters ϕ_1, \dots, ϕ_n , so that it is not possible to give a useful interpretation for the size of θ_T .

Siganos, 2008; Taylor and Braddock, 2007 for the “size bias”). The ambiguity in the interpretation of university size is probably one reason why some rankers publish distinct rankings of universities by size, for example in the case of the Italian ranking of universities (Censis and La Repubblica, 2012).

The fourth term, $T_{1,i}$, is a non-linear function of the parameters reflecting the university’s preferences over the different outputs that can be produced. In the model outlined above, a difference in preferences between two universities has an effect on the score which is independent of the effect of expenditure per student and of the size of the university. The same holds for the effect of a difference in input prices, which is given by the fifth right-hand term ($T_{2,i}$) of Equation (6).

The last term of Equation (6) gives the error term of the empirical specification. This, in turn, is equal to the difference between two weighted sums, where the weights are the parameters of the different outputs in the score function: v_i is the weighted sum of the measurement errors of the ranker; u_i is the weighted sum of the inefficiency terms from the system of Equations (2). Error in the measurement of university outputs is likely to occur, as even indicators based on apparently objective statistics can often be problematic (e.g. Enserink, 2007).

Conversely, we do not expect to observe productive inefficiency in our sample. Although higher education is mostly a not-for-profit industry (which is sometimes seen as a possible cause of inefficiency since market forces are not operating as in other industries), it has been exposed to increased competition in the last decades (Bonaccorsi and Daraio, 2007). Furthermore, our sample only includes universities which distinguished themselves in terms of the indicators collected by THES-QS. One reason for that could be that they have made efficient use of their resources. For example, in his assessment of efficiency in British higher education using data envelopment analysis, Johnes (2006) finds that 50 to 60% of British universities lie on the multiple output production frontier, displaying the maximum possible efficiency score. Even though the UK is very well represented in the ranking, “only” 32 institutions reached the top 200 according to THES-QS, fewer than the number of universities deemed efficient in Johnes’ (2006) analysis.

3. Data and methods

To estimate Equation (6), data at the institutional level is needed on: the dependent variable, which is the external assessment given to the institution by a ranker; the expenditure per student and the number of students; a variable representing differences in mission across institutions; and some

variables to capture the effects of different input prices and, possibly, other characteristics related to the geographic location. With this information, it is possible to estimate Equation (6) by OLS.

Concerning the external assessment, we use data from the 2007 THES-QS World University Ranking (O'Leary, Quacquarelli and Ince, 2008). The THES-QS ranking is an international ranking of universities published annually by the Times Higher Education Supplement. This ranking aims at being holistic, comprising the most relevant aspects on which universities can be evaluated (research, teaching, employability of students). The ranking is computed according to a score which, in turn, is the weighted sum of six standardised sub-scores. The six measures are: peer review (weight 0.4), recruiters review (weight 0.1), citations score (based on citations of published academic papers per faculty member - weight 0.2), faculty-students score (based on teaching staff / students ratio - weight 0.2), international students score (based on international students / total students ratio - weight 0.05), and international faculty score (international faculty / total faculty ratio - weight 0.05).

The ranking strongly relies on surveys of academic staff and of recruiters, whose combined weight is 50%. Taylor and Braddock (2007) and Marginson (2007) argue that such measures only replicate subjective opinions of academics or recruiters instead of generating a comparison based on objective information or value added. This is a common problem with reputational assessments, encountered in rankings as different as, for example, football rankings (e.g., Goff, 1996, and Lebovic and Siegelmann, 2001). Bowman and Bastedo (2010), using data from the THES-QS ranking, find evidence that past scores in the THES-QS ranking affect current scores by reinforcing the reputation of an institution. In contrast, Cyrenne and Grant (2009) find that the change in reputation in the Canadian McLean's ranking "is based on what might be considered the appropriate 'signals' that one might use in estimating the respective quality of the institution" (p. 237).

Another concern with the THES-QS is related to the findings of Saisana et al. (2011) who argue, on the basis of a robustness analysis of the THES-QS and other rankings, that these international rankings are not suitable for comparing the position of individual universities. The reason is that using the same sub-scores but changing the aggregation methodology, the position of individual universities can change substantially. However, we are not very concerned about this problem, since measurement error in the dependent variable is not a source of bias for an OLS estimate (see e.g. Greene, 2003, Chapter 12). Furthermore, the fact that the position of a university would change by changing the rankers' weights is predicted by our model, as it is the fact that the effect of expenditure should remain positive as long as the weights for the outputs are positive.

As our dependent variable, we use the log of the final score received by a university. We assume that the overall score can be considered as an approximation of a Cobb-Douglas function where several university outputs enter as the argument, as in Equation (3).

Our main explanatory variable is the logarithm of expenditure per student (*expenditure*). For measuring expenditure, we used the total consolidated operating expenditure for the chosen fiscal year (normally, 2006/2007 or 2007). This means that expenditures for hospitals or big laboratories or firms owned by universities are included in our estimates.³ This is likely to introduce some noise in the measurement of expenditure, which probably leads the coefficient in the OLS estimates to be biased towards zero. The expenditure has been adjusted for the Purchasing Power Parity (PPP) index. Student numbers have been collected from O’Leary et al. (2008). Their logarithm, *studs* is used in the estimations as an indicator of the size of the university. We use the total number of students, and not the full-time equivalent.

The term $T_{1,i}$, in Equation (6), is a function of γ_i , which represents the university’s preferences over the production of different outputs. These preferences are defined over a multiplicity of dimensions, and it would be very difficult to adequately represent all of them. However, we collected data on two indicators related to some of the dimensions constituting the mission of a university.⁴ To capture how research-oriented an institution is, we follow Bonaccorsi and Daraio (2007) and use the ratio between PhD students and total students at an institution (*ratio_{phd}*). More precisely, we use the number of PhD degrees awarded in the last 12 months as the numerator (Quacquarelli Sismonds, 2009). The underlying assumption is that a research-oriented university gives higher priority to PhD training. Arnold (2008) shows the synergy between teaching and research using data from the Erasmus University Rotterdam. This is more pronounced for advanced levels of education. As commonly done when working with shares in empirical applications, we use a logistic transformation of the share of PhD students, so that the variable that we use in our estimations (*phd*) is equal to $\ln(\text{ratio}_{phd} / (1 - \text{ratio}_{phd}))$. We present the results with and without the inclusion of this variable, since using it implies a substantial loss of observations (see Table 1).

³ This has been done for convenience’s sake: total consolidated expenditure can be found for universities around the world from websites, while “pure” (non-consolidated) expenditures for education and research can only be acquired by a detailed questionnaire (see e.g. Aghion et al., 2007, 2010; Daraio et al., 2011). We considered this unnecessary as it is unlikely that a university owns institutions completely unrelated to its research or educational goals.

⁴ The use of proxies for university mission differs from the approach followed by studies estimating cost functions of higher education institutions. The literature has dealt with mission diversity through parameter heterogeneity, either by estimating cost functions for different types of institutions separately (Johnes et al., 2004), or by allowing for different parameters for each institution (Johnes and Johnes, 2009).

According to the flexible specification of Equation (2), the output can also be distinguished by field of study (e.g., graduates in business studies or agriculture). For this reason, we construct an index of specialisation which can take a value between zero and six, reflecting the specialisation of universities in different research fields.⁵ This index is based on the field-specific reputation rankings published together with the THES-QS ranking. It takes higher values for universities ranking better in disciplines such as biology, physics, medicine or engineering than in the humanities or the social sciences. Although this measure is intended to capture the relative specialisation of a university in science and engineering, we cannot exclude that it is endogenously correlated with the peer review score, which is based on the field-specific rankings. Due to this concern, we regress our index of specialisation on the peer review score, and we generate a variable that is equal to the residuals of this regression, *hardscience*. This variable is uncorrelated by construction with the peer review score, but it is very highly correlated (0.99) with the original index of specialisation. It is not easy to form expectations for the sign of the coefficient of this variable. On one hand, the THES-QS ranking has sometimes been considered to be “biased” against universities in the humanities and social sciences (e.g. Baty, 2010; Holmes, 2006). On the other hand, it can be argued that the technology needed for research and learning is more expensive in the “hard sciences” than in other fields of study due to the need for complex machineries and laboratories. As a result, institutions focusing on the “hard sciences” will probably rank worse than other universities with the same output, conditional on their expenditure.

Differences in input prices can also affect the score. For this reason, expenditures in different countries are converted using the PPP index. However, producer prices faced in the higher education sector might evolve differently than other prices in the economy (see e.g. Commonfound Institute, 2013 for the U.S.). As a result, in one specification, we add dummy variables representing the geographic area of the institution for taking into account differences in prices that originate in a given geographic area. The dummy variables that we generated are *US*, *UK*, *Eu* (meaning all Western Europe but Great Britain), *Canada*, *Oceania* (Australia and New Zealand), *Asia* (Japan, Hong Kong, Singapore, Taiwan, South Korea) and *others* (Brazil, Mexico, South Africa and Israel). However, it must be noticed that the coefficients of these dummies will also reflect the effect of other unobserved factors operating at a geographic-area level. For example, these dummies could reflect

⁵ O’Leary et al. (2008) identifies five fields of study, and ranks universities according to each field, based on the peer review sub-score. These fields are technology, natural sciences, bio-medicine and life sciences, social sciences, and arts and humanities. We classify the first three of them as “hard fields”. The index of specialisation is generated by assigning one point for each possible couple comprising one “hard field” and one “soft field” where the hard field ranks better. Thus, the index reaches its maximum score (i.e., six) for a university ranking better in all three “hard fields” than in any of the “soft” ones. Conversely, a university ranked higher in social sciences and arts and humanities than in the other three fields receives zero points.

differences in mission among universities located in different geographic areas which are not accounted for by the share of PhD students.

In order to see whether the results are robust to taking unobserved heterogeneity into account, we also collected data on the THES-QS 2005 ranking. In 2005, THES-QS published the list of the top 500 universities in the world. As a result, we can associate every university in our sample with its position in the 2005 ranking. The overall score was published only for the top 200 universities, but the correlation between the log score and the log rank for the 2005 ranking is almost perfect (-0.99). For this reason, we predicted the log score for 2005 on the basis of the rank, and used it in our estimates as the lagged dependent variable (LDV).⁶

Wooldridge (2002, p. 66) offers an interesting interpretation of the coefficient of a variable in a cross-sectional regression with LDV, which we extend to our case. Given two universities with the same starting score in 2005, the coefficient gives the part of the difference in the 2007 score of the two universities that is associated to a one-percent variation in the regressor. The inclusion of an LDV allows to rule out reverse causality and to control for time-constant, unmeasured factors influencing both the independent and the dependent variables. Reverse causality could occur if a higher position in the ranking allowed universities to demand higher tuition fees, which in turn would raise the per student expenditure. For this reason, we think that it is useful to see, as a robustness check, if the results are qualitatively the same when including the LDV. However, it is well known (e.g. Achen, 2000) that the coefficients in a cross-sectional regression with the LDV are severely biased towards zero. Angrist and Pischke (2009, App. 5.4) formally address the problem and show that, if the fixed-effects model is correct (i.e., there is a time-constant, unobserved factor correlated with both dependent and independent variables), then the cross-sectional regression with the LDV provides a lower bound for the coefficients to be estimated.⁷

Table 1 presents the number of universities, by geographic area, contained in the entire population of the top 200 institutions in the sample that is available if we do not include *phd*, and in the sample used for our baseline estimation (comprising the observations for which data on the relative number of PhD students is available). The sample contains 184 observations of the 201 universities belonging to the list, but it includes only 156 observations if the variable *phd* is included in the analysis. Adding

⁶ The methodology of the THES-QS ranking has been subject to slight modifications between 2005 and 2007. For example, the way in which the opinions of the surveyed academics are aggregated into the peer review score has changed. However, we believe that the THES-QS score of 2005 still adequately captures those unobserved characteristics of a university that we wish to control for.

⁷ In the LDV regressions, we do not include the geographic dummies, because in that case the coefficient of a geographic dummy will reflect the change in the score common to universities in a given geographic area. This change, in turn, could be explained by other explanatory variables. As a result, the attenuation bias would be exacerbated by including the geographic dummies.

this variable substantially decreases the number of European and Canadian universities. The fourth row of Table 1 shows the numbers of universities in the top 200 according to the 2012/2013 THES ranking. The last row shows the numbers of institutions that are in the top 200 in 2012/2013 and for which we have data on our explanatory variables. This group of universities is the sample that we use to run the regressions whose results are reported in Column IX of Table 4 and Table 5. Based on these figures it is possible to see that between 2007 and 2012 the geographical composition of the ranking remained quite stable, although the presence of the US in the top 200 has increased at the expense of Asia, Oceania, and Canada.

Table 2 displays a number of descriptive statistics of our sample. The variability with respect to different characteristics within the sample is very high. Expenditures per student range from a minimum of 6,612€ (Università di Bologna) to an astonishing 861,243€ (Caltech)⁸, with a median of 26,324€. The number of students ranges from 2,197 (Caltech) to 155,392 (Universidad Nacional Autónoma de México), with a median equal to 19,130.

Table 1
Number of universities

	Asia	Canada	Eu	Oceania	UK	US	Others	total
Top 200, 2007	26	11	54	15	32	57	6	201
Top 200, 2007 (sample without <i>phd</i>)	19	11	47	15	32	56	4	184
Top 200, 2007 (sample with <i>phd</i>)	16	6	36	13	31	52	2	156
Top 200, 2012/2013 (all institutions)	18	8	53	9	31	76	5	200
Top 200, 2012/2013 (sample with <i>phd</i>)	10	3	28	8	26	52	1	128

Table 2
Descriptive statistics

	Min	Max	Average	Std. Dev.	Median
Students	2197	155392	23048	16779	19130
Exp. per student (€)	6612	861243	47288	75154	26324
Exp. per student (€ - US institutions)	13497	861243	99950	120130	64393
Exp. per student (US = 0)	6612	67525	24248	11441	21605
Phd ratio	0.0035	0.0851	0.0218	0.0133	0.0186

⁸ Caltech, which researches for the N.A.S.A., includes in its accounts the expenditures for the Jet Propulsion Laboratory (which, in turn, represents more than 75% of the total expenses reported), and has also by far the lowest number of students and the highest level of *relphd*. However, after the logarithmic transformation the observation does not appear as an outlier according to our diagnostic measure which (Wilks statistic). The results are not substantially affected by the presence of this institution.

As a diagnostic measure for outlying observations we use the Wilks statistic (Belsley et al., 2004, Chapter 2), computed on the variables included in our baseline estimation. This leads us to exclude one observation, the University of St. Louis in Washington (Wilks statistic = 0.15). More details are provided in Appendix B.

4. Results

4.1 Main specification

Table 3 presents the estimation results of our baseline regression. For every variable, we report the coefficient estimated by OLS with standard errors robust to heteroscedasticity. Given the limitations in our measures, the R-squared is surprisingly high: the model explains about 47% of the variation in the log score. Considering the substantial noise that is likely to characterise our measures, and the omission of the price level of inputs as a control variable, we consider this remarkably large. All estimated coefficients are significant at a 1% level. The coefficient of *expenditure* is positive, indicating an elasticity of the THES-QS score for expenditure per student of about 9%. The variable *phd* is also positive, indicating that mission matters and that research-oriented universities fare better in the ranking. Although the number of students (*studs*) also takes on a positive coefficient, it is not possible to say whether this is evidence of positive returns to scale in top universities or whether it is just a result of the weights applied by the 2007 THES-QS ranking. The index of specialisation in “hard sciences” takes on a negative sign. This suggests that it is more difficult for universities specialised in the hard sciences to score high on the assessment, conditional on their expenditure, size and research orientation.

Using the residuals from the estimation reported in Table 3, we test the null hypothesis of absence of inefficiency. Our test, which follows Coelli (1995), is a test of negative skewness of the estimated error term. This test is based on the idea that if the u_i component in Equation (6) is large compared to the v_i component, then the distribution of the error term estimated by OLS is negatively skewed. This test is based on the OLS estimates, and it does not require to estimate a stochastic frontier production function. The estimation results suggest that we cannot reject the hypothesis of absence of productive inefficiency in our sample. The p -value is very high (0.47), indicating that there is very little to be gained by modelling the error term differently than in OLS. In fact, we also performed a maximum-likelihood estimation of a stochastic production frontier model with half-normally distributed inefficiency disturbances (Aigner et al., 1977), finding that the differences to the OLS

estimates are negligible. The results from the stochastic frontier estimation are not reported in this paper.

Table 3

Baseline estimates of the effect of expenditure per student and other variables on the university's log THES-QS score in 2007

Expenditure	0.092***	(0.011)
Phd	0.139***	(0.019)
Studs	0.076***	(0.017)
Hardscience	-0.015***	(0.004)
Constant	3.73***	(0.181)
Observations	155	
R-squared	0.47	

** Significant at a 5% level

*** Significant at a 1% level

(Standard errors in brackets)

The elasticity of expenditure per student allows us to indicate how much the rank of a university would improve if a university were to increase its expenditures. To answer this question, it is useful to know that the rank is approximately linear in the log score: a regression of rank on log score yields an R-squared equal to 0.97, and the estimated coefficient of log score is -354. This implies that a 1% increase in expenditure per student translates into a predicted gain of approximately 0.32 positions in the ranking. As a rule of thumb, we can say that to gain one position in the ranking, a university needs to increase its expenditure by 3%.

4.2 Robustness checks

Table 4 reports our robustness checks. In the first column, we add the dummies for the geographic location of universities (the reference is continental Europe). The estimate of the elasticity of score for expenditure increases to 14%, and the coefficients for the geographic dummies for UK, Canada and Oceania are positive and significant, indicating that universities from those countries (or continents) fare better than European ones in the ranking. Interestingly, this is not the case for the US. In the second column, we omit the share of PhD students, which results in an increased sample size (the 184 observations described in the second row of Table 1, albeit excluding the outlier). The main difference to the baseline estimates is that the number of students and a specialisation in "hard sciences" are not significantly associated with the score, although the p -value for the number of students is close to 5% and the sign for the two coefficients remains unaltered.

To check whether the results are driven by US and British universities (which constitute more than half of the sample when the information on PhD students is used) we estimate the model separately for US-UK universities and for other universities. The two estimates are reported in Columns (III) and (IV) for US-UK and other countries, respectively. A Chow test (Greene, 2003, sec. 6.4.1) indicates that

running the two estimates separately fits the data significantly better (p -value=0.04) than running one single regression as in Table 3. However, the only substantially different coefficients between the estimates based on the two different samples are for *phd* and the constant term. In order to see whether this casual observation can be substantiated, we estimated a model in which the constant term and the coefficient for *phd* are different for US-UK universities. This regression is reported in column (V). We test this model against the baseline model in Table 3 using a test for exclusion restrictions (Greene, 2003, sec. 5.3.3), rejecting the null hypothesis that the two models are equivalent at a 1% confidence level. Next, we use a Chow test to test this model against a model in which the coefficient is allowed to be different for all variables in the model (i.e., columns III and IV combined). In this case, we fail to reject the null hypothesis that the two models are equivalent (p -value=0.43). These two tests confirm our observation that both the constant term and the coefficient for the share of PhD students differ between US-UK and other universities. Conversely, the coefficient of expenditure per student does not differ significantly in estimates based on the two different subsamples.

In column (VI) we report the estimation results obtained by excluding the top 20 universities in the ranking. The results are not substantially affected, although the coefficient of expenditure is lower than in the baseline model (0.055). In column (VII) we re-compute the score by excluding the faculty-student ratio, and we use its logarithm as the dependent variable. We do this because the faculty-student ratio can be considered an input rather than an output while the dependent variable is intended to be a measure of university output. As we computed the score by excluding the faculty-student score, the same can be done for the other five sub-scores. The results are very similar and have not been reported here for brevity, not reported here for the sake of brevity, are very similar to the baseline estimation. The reason is that, with the exception of the two indicators for internationalisation (which account for only 10% the score), all components of the score are positively and significantly related to expenditure per student.

Column (VIII) reports the results obtained when the outlying observation is included. In both cases, results are very similar to those of the baseline specification. In column (IX) we use the log score of the THES 2012/2013 ranking of universities as the dependent variable. The purpose is to see whether introducing a time lag between the dependent and the independent variables would substantially change our results. This does not seem to be the case.⁹ In column (X) we use a different measure of size, which is faculty staff. Hence, expenditure is divided by the number of faculty staff. In addition,

⁹ The number of observations is only 127 in this regression, because we observe only institutions for which we collected data in 2009 and which subsequently appeared in the 2012/2013 ranking. Notice that the 2012/2013

Table 4
Estimates of the effect of expenditure per student and other variables on the ranking score – Robustness checks

	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)
Expenditure	0.141*** (0.023)	0.117*** (0.012)	0.068*** (0.015)	0.109*** (0.037)	0.075*** (0.014)	0.055*** (0.012)	0.095*** (0.016)	0.089*** (0.012)	0.139*** (0.017)	0.077*** (0.013)
Phd	0.139*** (0.02)		0.215*** (0.027)	0.085*** (0.025)	0.098*** (0.022)	0.101*** (0.018)	0.112*** (0.027)	0.129*** (0.021)	0.175*** (0.03)	0.139*** (0.020)
Studs	0.104*** (0.019)	0.032 (0.017)	0.081*** (0.024)	0.09*** (0.028)	0.081*** (0.018)	0.07*** (0.017)	0.118*** (0.023)	0.074*** (0.017)	0.13*** (0.025)	0.095*** (0.017)
Hardscience	-0.012*** (0.004)	-0.005 (0.005)	-0.016*** (0.005)	-0.012 (0.007)	-0.013*** (0.004)	-0.01** (0.005)	-0.016*** (0.006)	-0.013*** (0.005)	-0.015** (0.007)	-0.013*** (0.005)
US	-0.026 (0.001)									
UK	0.117*** (0.001)									
Asia	-0.011 (0.001)									
Canada	0.133*** (0.001)									
Oceania	0.176*** (0.001)									
Others	-0.014 (0.001)									
US-UK *					0.11*** (0.032)					
phd					0.471*** (0.105)					
US-UK										
Constant	3.255*** (0.247)	3.529*** (0.188)	4.077*** (0.278)	3.309*** (0.323)	3.551*** (0.192)	3.748*** (0.165)	2.979*** (0.246)	3.724*** (0.179)	3.044*** (0.274)	3.64*** (0.166)
Obs	155	183	82	73	155	136	155	156	127	155
R-squared	0.604	0.282	0.567	0.276	0.513	0.246	0.315	0.432	0.502	0.433

score is computed according to a very different methodology than the 2007 score. As a result, the coefficients from Model 11 are not entirely comparable to the coefficients from other models.

Table 5
Estimates of the effect of expenditure per student and other variables on the ranking score, conditional on lagged score

	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)
Expenditure	0.039*** (0.009)	0.042*** (0.009)	0.026** (0.011)	0.049** (0.023)	0.026** (0.01)	0.037*** (0.011)	0.029** (0.012)	0.035*** (0.01)	0.077*** (0.015)	0.042*** (0.011)
Phd	0.05*** (0.016)	0.015 (0.01)	0.064** (0.027)	0.043** (0.02)	0.055*** (0.018)	0.049** (0.019)	0.012 (0.027)	0.039** (0.018)	0.037 (0.025)	0.079*** (0.020)
Studs	0.029** (0.012)	-0.004 (0.003)	0.03** (0.015)	0.029 (0.024)	-0.01 (0.025)	0.018 (0.017)	0.039** (0.019)	0.025** (0.012)	0.077*** (0.019)	0.053*** (0.015)
Hardscience	-0.006 (0.003)	0.323*** (0.023)	-0.004 (0.005)	-0.008 (0.005)	0.028** (0.012)	-0.006 (0.004)	-0.008 (0.005)	-0.004 (0.004)	-0.001 (0.005)	-0.008** (0.004)
Score05	0.281*** (0.027)		0.254*** (0.035)	0.347*** (0.043)	-0.005 (0.003)	0.297*** (0.041)	0.323*** (0.03)	0.29*** (0.028)	0.319*** (0.033)	0.371*** (0.047)
US-UK *					0.286*** (0.029)					
phd										
US-UK					-0.003 (0.103)					
Constant	3.07*** (0.143)	2.855*** (0.12)	3.272*** (0.207)	2.772*** (0.248)	3.113*** (0.149)	3.124*** (0.166)	2.678*** (0.216)	3.043*** (0.146)	2.15*** (0.214)	2.375*** (0.228)
Obs	155	183	82	73	155	104	123	156	127	155
R-squared	0.726	0.706	0.764	0.625	0.735	0.513	0.673	0.708	0.714	0.645

instead of the number of students, we use the number of faculty staff as a control variable (the resulting coefficient is indicated in the same row as the coefficient for student numbers in the other regressions). The results are very similar to those of the baseline specification.

To see whether our estimation results suffer from reverse causality or unobserved heterogeneity, Table 5 reports the LDV estimates. Column (I) of Table 5 reports the results for the regression presented in Table 3. However, in this iteration the THES-QS 2005 score appears as an independent variable. Column (I) of Table 5 can also be compared to Column (I) of Table 4, with the 2005 score replacing the geographic dummies. All other columns of Table 5 report the results for the regression in the same column as in Table 4, albeit with the 2005 score included as a dependent variable. The coefficient for expenditure per student is constantly positive and significant, although its size is consistently smaller than in Table 4. It is equal to 0.04 in Column (I), compared to 0.09 in Table 3. This was expected, since the coefficient probably suffers from a considerable attenuation bias. The coefficients for the number of students and the PhD ratio are consistently positive, but lower than in Table 4. They are significant only in some of the specifications. A focus on hard sciences seems to affect the 2005 score negatively, but this result is not significant. The coefficient for the lagged dependent variable ranges from a minimum of 0.25 in Column (III) to a maximum of 0.35 in Column (IV). Again, it is possible to compare the fit of the model with all observations in Column (I) with the model separately estimated for US-UK and other universities (Columns III and IV) using a Chow test. The p-value for the test is 0.10, indicating that that the coefficients may be the same for both samples. Finally, we notice that the fit of the estimates improves considerably as a result of the inclusion of the lagged dependent variable: for example, the R-squared is 0.73 in Column (I), whereas it is 0.47 in Table 3.

A number of additional checks have been performed, albeit these are not reported in the paper. For example, we included a dummy variable which is equal to 1 if a university owns a hospital, since the costs for the hospital would appear in the total consolidated expenditure of the university. We also included a dummy for public or private ownership. However, the coding of both variables requires considerable judgement on the side of the researcher, because the relationship between universities, governments and academic hospitals can be complex. Furthermore, the results indicated that these variables are not significant predictors of the ranking score. Hence we do not report these estimates. We also included interaction terms between expenditure per student, the PhDs-to-students ratio and the number of students. When including the interaction terms, the key results of the paper remain unchanged: all three variables (expenditure per student, PhDs-to-students ratio and number of students) are significant predictors of the score. However, the interaction terms and the original variables are highly collinear. As a result, the coefficients and significance of the variables depend on

which specification is chosen (whether outliers are excluded, dummies are included, etc.). For this reason, we do not report the results.

5. Conclusions

In this paper, we analyse the relationship between a university's expenditure per student and its score in the 2007 THES-QS ranking. We develop a model in which a university produces multiple outputs using multiple inputs, and a ranking institution assesses the university. The model shows that expenditure per student is positively related to a university's score. This holds true even if the ranker weights the university outputs differently than the university, and if there is random noise in the ranker's measures of output. The result depends on the assumption that both the university utility and the expected value of the ranker assessment are non-decreasing functions of the university's outputs. The score of a university is affected by other important factors, such as the mission of a university and inefficiency in production.

The estimated coefficients of the variables included in the model are, in general, significant and with the expected sign. The model explains a substantial part of the variance of the dependent variable (47% in our baseline specification). Expenditure per student seems to increase the score of a university, with an elasticity of 9% in our baseline specification. This means that a university increasing the expenditure per student by 3% would gain one position in the THES-QS ranking on average.

The finding of a positive elasticity of the university score in the 2007 THES-QS ranking for the expenditure per student is robust to a big number of changes in the empirical specification, some of which have been reported in this paper. Importantly, the coefficient of expenditure per student remains positive and significant after controlling for the past score in the THES-QS ranking.

Research-oriented universities seem to be at an advantage in the ranking, confirming that the mission of a university is an important determinant of the score. Larger institutions rank better than smaller ones, although the interpretation of this finding is not clear in light of the model. Evidence is less clear for being oriented towards "hard sciences" (life and natural sciences and engineering). If anything, this seems to put a university at a disadvantage. Universities from the UK, Oceania and Canada fare relatively well in the ranking, but US universities apparently do not, given their level of expenditure per student.

References

- Achen, C. H. (2000). *Why Lagged Dependent Variables Can Suppress the Explanatory Power of Other Independent Variables*.
- Aghion, P., Dewatripont, M., Hoxby, C., Mas-Colell, A., and Sapir, A. (2007). *Why Reform Europe's Universities?* (No. 04) (pp. 1–8).
- Aghion, P., Dewatripont, M., Hoxby, C., Mas-Colell, A., and Sapir, A. (2010). The Governance and Performance of Universities: Evidence from Europe and the US. *Economic Policy*, (January), 7–59.
- Aigner, D., Lovell, C. A. K., and Schmidt, P. (1977). Formulation and Estimation of Stochastic Frontier Production Models. *Journal of Econometrics*, 6, 21–37.
- Angrist, J. D., and Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion. An empiricist's companion* (p. 392). Princeton: Princeton University Press. doi:10.1057/be.2009.37
- Arnold, I. J. M. (2008). Course Level and the Relationship between Research Productivity and Teaching Effectiveness. *The Journal of Economic Education*, 39(4), 307–321. doi:10.3200/JECE.39.4.307-321
- Baty, P. (2010). Ranking Confession. *Inside Higher Ed*. Retrieved July 18, 2013, from <http://www.insidehighered.com/views/2010/03/15/baty>
- Belsley, D. A., Kuh, E., and Welsch, R. E. (2004). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity* (2nd ed.). Hoboken: John Wiley & Sons.
- Bijwaard, G. E. (2010). Immigrant migration dynamics model for The Netherlands. *Journal of Population Economics*. doi:10.1007/s00148-008-0228-1
- Bonaccorsi, A., and Daraio, C. (2007). Theoretical Perspectives on University Strategy. In A. Bonaccorsi & C. Daraio (Eds.), *Universities and Strategic Knowledge Creation* (pp. 3–30).
- Bonaccorsi, A., Daraio, C., Rätty, T., and Simar, L. (2007). *Efficiency and University Size: Discipline-wise Evidence from European Universities* (No. 10265). Retrieved from <http://mpra.ub.uni-muenchen.de/10265/>
- Bowman, N. a., and Bastedo, M. N. (2010). Anchoring effects in world university rankings: exploring biases in reputation scores. *Higher Education*, 61(4), 431–444. doi:10.1007/s10734-010-9339-1
- Censis, and La Repubblica. (2012). *La Grande Guida all'Università*. Roma: Gruppo Editoriale L'Espresso.
- Coelli, T. I. M. (1995). Estimators and Hypothesis Tests for a Stochastic Frontier Function: A Monte Carlo Analysis. *The Journal of Productivity Analysis*, 6, 247–268.
- Commonfund Institute. (2013). *Higher Education Price Index* (pp. 0–23). Retrieved from www.commonfund.org

- Cyrenne, P., and Grant, H. (2009). University Decision Making and Prestige: An Empirical Study. *Economics of Education Review*, 28(2), 237–248. doi:10.1016/j.econedurev.2008.06.001
- Daraio, C., Bonaccorsi, A., Geuna, A., Lepori, B., Bach, L., Bogetoft, P., ... Eeckaut, P. Vanden. (2011). The European University Landscape: A Micro Characterization Based on Evidence from the Aquameth Project. *Research Policy*, 40(1), 148–164. doi:10.1016/j.respol.2010.10.009
- Dichev, I. D. (1999). How Good Are Business School Rankings? *The Journal of Business*, 72(2), 201–213.
- Enserink, M. (2007). Who Ranks the University Rankers? *Science*, 317(5841), 1026–1028.
- Fernández, C., Koop, G., and Steel, M. (2000). A Bayesian Analysis of Multiple-Output Production Frontiers. *Journal of Econometrics*, 98, 47–79.
- Fernández, C., Koop, G., and Steel, M. F. J. (2002). Multiple-Output Production with Undesirable Outputs: An Application to Nitrogen Surplus in Agriculture. *Journal of the American Statistical Association*, 97(458), 432–442.
- Fried, H. O., Lovell, C. A. K., and Schmidt, S. S. (2008). Efficiency and Productivity. In H. O. Fried, K. C. A. Lovell, & S. S. Schmidt (Eds.), *The Measurement of Productive Efficiency and Productivity Growth* (pp. 1–106). New York: Oxford University Press.
- Goff, B. (1996). An Assessment of Path Dependence in Collective Decisions. Evidence from Football Polls. *Applied Economics*, 28, 291–297.
- Greene, W. H. . (2003). *Econometric Analysis*. Prentice Hall (Vol. 97, p. 1026). doi:10.1198/jasa.2002.s458
- Hanushek, E. a., and Woessmann, L. (2012). Do Better Schools Lead to more Growth? Cognitive Skills, Economic Outcomes, and Causation. *Journal of Economic Growth*, 17(4), 267–321. doi:10.1007/s10887-012-9081-x
- Hazelkorn, E. (2008). Learning to Live with League Tables and Ranking: The Experience of Institutional Leaders. *Higher Education Policy*, 21(2), 193–215. doi:10.1057/hep.2008.1
- Holmes, R. (2006). The THES University Rankings: Are They Really World Class? *Asian Journal of University Education*, 1(1), 1–14.
- Ince, M. (2007). *World University Rankings 2007*.
- Izadi, H., Johnes, G., Oskrochi, R., and Crouchley, R. (2002). Stochastic Frontier Estimation of a CES Cost Function: The Case of Higher Education in Britain. *Economics of Education Review*, 21, 63–71.
- Jehle, G. A., and Reny, P. J. (2001). *Advanced Microeconomic Theory*. *Annals of Physics* (Second Edi., Vol. 2). Addison Wesley Longman.
- Johnes, G., and Johnes, J. (2009). Higher education institutions' costs and efficiency: Taking the decomposition a further step. *Economics of Education Review*, 28(1), 107–113. doi:10.1016/j.econedurev.2008.02.001

- Johnes, G., Johnes, J., Thanassoulis, E., Lenton, P., and Emrouznejad, A. (2004). *An Exploratory Analysis of the Cost Structure of Higher Education in England*. London.
- Johnes, J. (2006). Data envelopment analysis and its application to the measurement of efficiency in higher education. *Economics of Education Review*, 25(3), 273–288. doi:10.1016/j.econedurev.2005.02.005
- Kuo, J.-S., and Ho, Y.-C. (2008). The cost efficiency impact of the university operation fund on public universities in Taiwan. *Economics of Education Review*, 27(5), 603–612. doi:10.1016/j.econedurev.2007.06.003
- Lebovic, J. H., and Sigelman, L. (2001). The forecasting accuracy and determinants of football rankings. *International Journal of Forecasting*, 17(1), 105–120. doi:10.1016/S0169-2070(00)00064-9
- Lenton, P. (2008). The cost structure of higher education in further education colleges in England. *Economics of Education Review*, 27(4), 471–482. doi:10.1016/j.econedurev.2007.05.003
- Longlong, H., Fengliang, L., and Weifang, M. (2009). Multi-product total cost functions for higher education: The case of Chinese research universities. *Economics of Education Review*, 28(4), 505–511. doi:10.1016/j.econedurev.2008.11.002
- Marconi, G. (2013). Rankings, Accreditations, and International Exchange Students. *IZA Journal of European Labor Studies*, 2(1), 5. doi:10.1186/2193-9012-2-5
- Marginson, S. (2007). Global University Rankings: Implications in general and for Australia. *Journal of Higher Education Policy and Management*, 29(2), 131–142. doi:10.1080/13600800701351660
- Monks, J., and Ehrenberg, R. G. (1999). U.S. News & World Report's College Rankings: Why They Do Matter. *Change: The Magazine of Higher Learning*, 31(6), 42–51. doi:10.1080/00091389909604232
- O'Leary, J., Quacquarelli, N., and Ince, M. (2008). *Top Universities Guide* (2nd ed.). London: QS Quacquarelli Simonds.
- OECD. (2008). *Education at a Glance 2008*. Paris.
- Quacquarelli Simonds. (2009). Online Database. Retrieved from www.topuniversities.com
- Ritzen, J. M. M. (2010). *A Chance for European Universities*. Amsterdam: Amsterdam University Press.
- Saisana, M., d'Hombres, B., and Saltelli, A. (2011). Ricketty numbers: Volatility of university rankings and policy implications. *Research Policy*, 40(1), 165–177. doi:10.1016/j.respol.2010.09.003
- Sauder, M., and Lancaster, R. (2006). Do Rankings Matter? The Effects of U.S. News & World Report Rankings on the Admissions Process of Law Schools. *Law & Society Review*, 40(1), 105–134. doi:10.1111/j.1540-5893.2006.00261.x
- Siganos, A. (2008). Rankings, Governance, and Attractiveness of Higher Education: The New French Context. *Higher Education in Europe*, 33(2-3), 311–316. doi:10.1080/03797720802254205

Taylor, P., and Braddock, R. (2007). International University Ranking Systems and the Idea of University Excellence. *Journal of Higher Education Policy and Management*, 29(3), 245–260. doi:10.1080/13600800701457855

Wooldridge, J. D. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.

Appendix A

Consider the following problem:

$$\max_{x_{i1}, \dots, x_{im}, D_i} O_{i1}^{\gamma_{i1}} \dots O_{in}^{\gamma_{in}} D_i^{-\delta}$$

Subject to:

$$\begin{cases} O_{i1} = A_1 x_{i1}^{\alpha_{11}} \dots x_{im}^{\alpha_{1m}} e^{-u_{i1}} \\ \dots \\ O_{in} = A_n x_{i1}^{\alpha_{n1}} \dots x_{im}^{\alpha_{nm}} e^{-u_{in}} \end{cases}$$

$$B_i \geq \sum_{k=1}^m p_k x_{ik}$$

$$B_i = f + bD_i$$

This is a production problem very similar to the one that we discuss in Section 2. Production is the same as in Equation (2). However, the utility of the university is modified so that the number of students, D_i , also plays a role. Notice that δ can be interpreted as the sum of a set of parameters $\delta_1, \dots, \delta_n$, in the same way as θ . The budget of the university depends on the number of students in a linear way, as it would be if every student pays the same tuition fee, or the university receives a constant, per-student subsidy from the government.

It can be verified that the optimal level of the choice variables $(x_{i1}, \dots, x_{im}, D_i)$ is:

$$x_{ik} = \frac{B_i \sum_{j=1}^n \alpha_{jk} \gamma_{ij}}{p_k \sum_{k=1}^m \sum_{j=1}^n \alpha_{jk} \gamma_{ij}} \quad \forall k$$

$$D_i = \frac{f \delta}{b(\sum_{k=1}^m \sum_{j=1}^n \alpha_{jk} \gamma_{ij} - \delta)}$$

In other words, the expenditure on every production input x_{i1}, \dots, x_{im} remains a constant fraction of the total expenditure.

As a result, it is again possible to substitute the optimal level of inputs into system of Equations (2) and then into (3), taking logs and manipulating, in order to arrive at Equation (6).

Appendix B

Figure 1 reports the Wilks statistic for every observation, ordered from the closest to 1 (most unproblematic) to the furthest away (most problematic). It is clear from the figure that there is one very problematic observation, which is the University of St. Louis in Washington (with a Wilks statistic of 0.15). Looking at different diagnostic measures (for example, the standardised residual from the regression, see Belsley et al., 2004, Chapter 1) leads to the same conclusion.

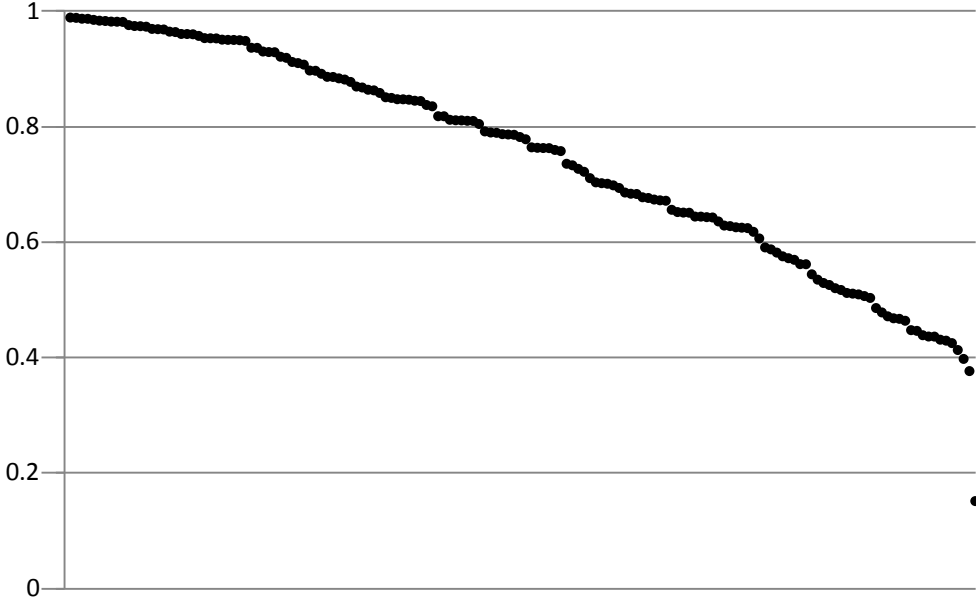


Figure 1 The Wilks statistic, from the closest-to-one (to the left) to the furthest observation (to the right).