Journal of Informetrics 11 (2017) 859–872

Contents lists available at ScienceDirect

Journal of Informetrics

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

Regular article

What drives university research performance? An analysis using the CWTS Leiden Ranking data

Koen Frenken, Gaston J. Heimeriks*, Jarno Hoekman

Innovation Studies, Copernicus Institute of Sustainable Development, Utrecht University, PO Box 80115, 3508TC, Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 20 May 2017
Received in revised form 13 June 2017
Accepted 21 June 2017

Keywords:
University ranking
Indicators
Excellence
Internationalization
Innovation
Country differences

ABSTRACT

This paper analyzes the factors underlying university research performance as indicated by the number of highly-cited publications, international co-publications, and university-industry co-publications. The three performance indicators evaluate three possible university missions, respectively: research excellence, internationalization, and innovation. Using a regression analysis, we assess to what extent a university's research performance is influenced by structural variables including size, age, city size, location in a capital city, disciplinary orientation, and country location. Our results show that research performance differences among universities mainly stem from size, disciplinary orientation and country location. This suggests that simple global benchmarking can be misleading; rather, benchmarking is most meaningful between universities of a similar size supplemented with contextual information on a university's specific mission, orientation and national institutions.

1. Introduction

Nowadays, we can find a ranking for almost every form of human activity. Academic activities at universities are no exception. We could, however, ask ourselves how meaningful such university rankings are and whether they are currently being used in a biased or naïve manner. Undeniably, there are large differences in performance among universities. Thus, the logic of university ranking seems appropriate: rankings reveal a university's performance compared to others. Indeed, university rankings are now proliferating. Apart from the most well-known ones such as the ARWU ("Shanghai") ranking, the Times Higher Education (THE) ranking and Quacquarelli Symonds (QS) ranking, at least 30 other rankings exist (Shin, Toutkoushian, & Teichler, 2011).

People both criticize and applaud global university rankings. While there is disagreement on which data, methodology and interpretations are the most robust (see for example Moed, 2017), many observers believe that global university rankings are here to stay. With students and academics facing greater options and opportunities, the existence of these rankings has heightened competition the world over and governments are now paying closer attention, even utilizing rankings to determine policies. Rankings are "performative" (Dahler-Larsen, 2011) in the sense that students, university boards and governmental bodies consider them meaningful, and rankings therefore influence their opinions, decisions, and actions. In most rankings, the aim is to compare so-called world-class universities, especially research-intensive ones. In doing so,

* Corresponding author.
E-mail addresses: k.frenken@uu.nl (K. Frenken), g.j.heimeriks@uu.nl (G.J. Heimeriks), j.hoekman@uu.nl (J. Hoekman).

http://dx.doi.org/10.1016/j.joi.2017.06.006
1751-1577/

rankings are creating a framework in which universities are part of a global knowledge system where global norms apply about what is considered performance.

Given the increasing impact of university rankings, and their questionable implicit assumption of a single global system, it has become pivotal to develop more reflexive and useful ways to interpret the results of rankings. Global university rankings suggest there is a single 'global' academic system with homogeneous university structures and objectives. Such a viewpoint ignores strong specificities among disciplines, countries, and university missions. Hence, university rankings have been criticized for misleading their users such as international students, job-hopping scientists, and policy officers (European Journal of Education, 2014; Kehm & Stensaker, 2009; Shin et al., 2011).

The fundamental problem underlying global university rankings is well articulated by Shin and Toutkoushian (2011, p. 2), who argued that "ranking universities is a challenging task because each institution has its own particular mission, focus and can offer different academic programs. Institutions can also differ in size and have varying amounts of resources at their disposal. In addition, each country has its own history and higher education system which can impact the structure of their colleges and universities and how they compare to others. It is therefore very difficult to rank entire universities, especially across national borders, according to the single criterion of ranking indicators". In our view, what is needed is an understanding of – and reflection on – the factors underlying university research performance. We will look at age, size, city size, capital city, disciplinary orientation and country location. An empirical analysis of university performance will give us insight into the structural differences among the best research universities worldwide. Consequently, we can form more specific – and more meaningful – peer groups that are relevant for benchmarking universities. For example, if size turns out to be very predictive of a high ranking – all other things being equal – a small university may compare itself more meaningfully with fellow smaller universities rather than those of any size. Or, if medical schools systematically perform differently compared to generic universities, we could argue that such schools should consider each other as relevant peers, rather than all types of universities. Furthermore, regression analysis can assess, for an individual university, whether its performance is better or worse than the expected value we can obtain from the regression coefficients. If the residual of an observation is positive (negative), this means the university in question is doing better (worse) than could have been anticipated from its structural features.

Our empirical study aims to analyze the factors underlying university research performance. We do so by using regression analysis to explain a university's performance from underlying structural variables. As dependent variables, we use three indicators provided by the CWTS Leiden University ranking, which has detailed bibliographical information on 750 universities worldwide for the period 2010–2013. The performance indicators we analyze are as follows: number of highly-cited publications, number of international co-publications, and number of co-publications with industry. The three indicators denote the evaluation criteria of three potential and different university missions. Highly-cited publications indicate research excellence, which many universities see as their goal. In addition to excellence, a second mission that some universities embrace is to act as an absorber of global knowledge through international networks. This goal is often pursued by universities in developing countries. Finally, universities can play a key role as sources of human capital and innovation for the economy. Some universities consider it as their main mission to contribute to the local economy, especially universities in peripheral regions (Bonaccorsi, 2016). In this paper, we use the number of highly-cited publications as a criterion of excellence, the number of international co-publications as a criterion of internationalization, and the number of university-industry co-publications as a criterion of a university's contribution to innovation.

Our study follows on previous studies that attempted to unravel the drivers of universities' research performance to advance a more careful use of ranking data. A study by Li, Shankar and Tank (2011) focused on national differences by taking for each country the number of universities reported in the ARWU 2008-ranking as dependent variable. The analysis, covering 93 countries, showed that – apart from population – GDP per capita, R&D expenditures and English as a language all contributed to the number of universities in the ARWU list. The residual analysis further revealed that UK and China are over-performers and the US an under-performer. A second study by Bornmann, Mutz and Daniel (2013) applied a multi-level analysis to analyze to what extent university output and country variables (population, GDP per capita, notably) affect the probability that a paper is among the top ten percent highest cited publications. They used the Leiden Ranking data for the period 2005–2009, which at the time was available for 500 universities. Their key result was that country variables explain the larger part of performance differences between universities (about eighty percent), while differences among universities matter relatively little (about twenty percent). The study was followed up by Bornmann, Stefaner, de Moya Anegón and Mutz (2014) applying the same multi-level approach on Scopus data while distinguishing between subject areas.

2. Leiden ranking

Although specific countries have been creating university rankings for certain disciplines for almost a century, global university rankings that cover many more universities and sciences are a relatively recent phenomenon (Kehm & Stensaker, 2009; Shin et al., 2011). Among the first was the *Academic Ranking of World Universities* (initially under the label of the Shanghai Ranking), published since 2003 by Shanghai Jiao Tong University in China. This ranking was soon followed by the 2004 edition of the THE Ranking achieved in partnership with QS.[1] Though these global rankings vary in their exact methodology, they are all based on composite indicators that combine different metrics in a single value by applying a weighting scheme. The disadvantage of these rankings is that weights are necessarily arbitrary; the results can be very sensitive to the choice of weights and the inclusion or exclusion of certain variables or the size of a university and its subject mix (Olcay & Bulu, 2016; Piro & Sivertsen, 2016; Shin et al., 2011). Some have tried to overcome these limitations by conducting multi-input, multi-output activities (Daraio, Bonaccorsi, & Simar, 2015).

The CWTS Leiden Ranking system, which was introduced more recently (Waltman et al., 2012), has been available from www.leidenranking.com since 2011.[2] This ranking differs in that it is based entirely on bibliometric indicators (from the Web of Science), in other words attributes of publications that can be considered more accurate than the information self-reported by universities or survey-based data. Nevertheless, Web of Science data has several well-known limitations including a bias towards English language journals (Van Raan, Van Leeuwen, & Visser, 2011) and the underrepresentation of applied studies, Arts and Humanities, and interdisciplinary research (Rafols, Leydesdorff, O'Hare, Nightingale, & Stirling, 2012). The Leiden Ranking contains uni-dimensional rather than composite indicators as in the ARWU, THE, and QS rankings which allows for the assessment of university research performance based on separate indicators instead of composite. Another advantage of the Leiden Ranking data is that, for the citation variables, it applies a normalization scheme per field to avoid biases resulting from varying citation practices across fields (Zitt, Ramanana-Rahary, & Bassecoulard, 2005).

What also makes the Leiden database appropriate for our analysis is its relatively wide coverage. We use the data from 750 universities for the period 2010–2013 from the 2015 CWTS Leiden Ranking (the latest version available at the start of this study), covering the 750 largest universities in the world in terms of publication output. Using mean values for a four-year period ensures that annual outliers have little influence on the results. The focus on the largest research universities avoids sampling on the dependent variable in our study (i.e. size is an independent variable). This also implies that we focus on research-intensive universities. Since such universities are concentrated in European countries and the US, our sample of universities is biased towards these countries.

The dependent variables are indicators directly drawn from the Leiden database and cover the identified university missions (variable name in brackets):

1. "*P_top10*": the number of publications that belong to the top 10% highest cited publications (*CITATION*)
2. "*P_int_collab*": the number of international co-publications (*INTERNATIONAL*)
3. "*P_UI_collab*": the number of co-publications with industry (*INDUSTRY*)

Throughout the study, we apply the Leiden data based on full counting, which gives equal weight (i.e. of one) to all publications of a university irrespective of the number of addresses on the publication.

Providing an analysis for these three measurements separately acknowledges that universities may have different missions. While universities all engage in teaching, some see research excellence as a second main priority (*CITATION*), while other universities focus more on collaboration abroad (*INTERNATIONAL*) or with firms (*INDUSTRY*) (Etzkowitz, Webster, Gebhardt, & Terra, 2000; Vorley & Nelles, 2008). Importantly, we do not claim that each of the three indicators is a universal and objective measurement of university performance *per se*. We prefer to consider each indicator more or less relevant for every single university, depending on its mission.

3. Data and methodology

We aim to explain the differences between universities for all three performance dimensions by observing various structural factors often claimed to affect university rankings (for a review and discussion, see Safon, 2013). *Firstly*, we take into account *SIZE* and *AGE*. We can expect that larger universities outperform smaller universities, even after correcting for size. Larger universities benefit from scale in that they can employ more sophisticated research equipment and run more specialist graduate programs. Scale economies may also exist when collaborating abroad (thanks to specialized international exchange programs as well as visibility), and when collaborating with industry (because of co-investments in specialized laboratories as well as in administrative overheads). Previous research has already shown that the citation impact of universities scales super-linearly with size (Van Raan, 2013). Thus, with each doubling of a university's output, the number of citations it receives more than doubles. In accordance with Van Raan (2013), we analyze the effect of university size by testing a scaling law where a scaling coefficient larger than one indicates scale advantages (see below). In doing so, we take the number of publications as the size variable using the full counting method. The effect of age on university performance is less well known. Yet the fact that some rankings provide separate results for young universities merely suggests that younger universities may be less likely to perform well than older universities, *ceteris paribus*. Indeed, as reputation may well impact the allocation of funding and mobility decisions – and reputation being partly affected by sheer age – we can

[1] In 2010, Times Higher Education and QS both started a separate ranking under the labels THE World University Rankings and QS World University Rankings, respectively.

[2] Similarly, the SCImago Institutions Ranking (www.scimagoir.com) is based on Elsevier's Scopus data; it is comparable with the Leiden Ranking but covers many more research institutions.

expect older universities to outperform younger ones. We measure age as 2015 minus the founding year of the university, which we obtained from the university websites. In the few cases where this information was not provided, we obtained the founding date first through Wikipedia and then through alternative Internet sources. Where there were mergers, we used the founding year of the oldest university.

Secondly, apart from the size and age of the university itself, we look at the metropolitan population of the city in which a university is located (*METROPOP*). Universities in larger cities may benefit from being co-located with other universities by sharing resources and combining teaching, as well as from their co-location with public research organizations or corporate R&D labs. Population data refers to the metropolitan area of the city and comes from the UNESCO Institute for Statistics (http://data.uis.unesco.org/). We also created a dummy variable where the university is located in a capital city (*CAPITAL*). This may be indicative for better access to public funding as well as a proxy for the overrepresentation of public labs.

Thirdly, we study disciplinary differences. Different fields of knowledge production have very different characteristics in their publication patterns, their citation patterns, and their engagement with students, industry, and other stakeholders (Heimeriks, 2013; Safon, 2013). We include dummy variables for technical universities (*TECHNICAL*) and medical schools (*MEDICAL*) as two internationally institutionalized types of university, with the generic university as the reference group, which generally covers science, medicine, social science, humanities (and in rarer cases also has an engineering faculty). We allocated the *TECHNICAL* and *MEDICAL* labels based on whether university names indicated either a technical university or medical school. For the remaining universities, we checked their homepages to see whether a university mentions a technical or medical specialization. We decided not to develop more detailed disciplinary characterizations of universities because the citation data, particularly known for its disciplinary sensitivity, has already been field-normalized in the CWTS Leiden Ranking dataset.[3]

Finally, we use country dummies (*COUNTRY*) to control for structural differences in national science systems and macro-conditions (Bornmann, Stefaner, de Moya Anegón, & Mutz, 2014). We expect countries to have structurally different scores in any performance ranking, mainly because the amount of funding and how it is allocated differs markedly between countries. While research is usually concentrated in universities, some countries have dedicated national research institutes. One prominent example is Germany, where top research is carried out at Max Planck institutes, that collaborate with universities but do not fall under any university. This alone may be a key explanation why German universities tend to rank lower than for example their British counterparts. Furthermore, there are institutional differences between universities related to medical research conducted at a medical faculty or within an academic hospital not included in the ranking. Below, we present 26 dummy variables referring to every country with more than five universities in the Leiden Ranking. All the other countries are therefore the reference group.

3.1. Model

We apply a log–log specification to identify the scaling coefficient vis-a-vis the size of a university as measured by the number of publications (Nomaler, Frenken, & Heimeriks, 2014; Van Raan, 2013). A scaling law describes one quantity (*Y*) as a function of the size of another quantity (*SIZE*), such that:

$$Y_i = \alpha SIZE_i^{\beta} \tag{1}$$

Here, we use the total number of publications of a university i (*SIZE*) as our size-variable. Empirically, taking natural logarithms on both sides of the equation and adding a random error term, the multiplicative form of Eq. (1) can be converted into a testable Eq. (2):

$$\ln Y_i = \ln\alpha + \beta \ln SIZE_i + \varepsilon_i \tag{2}$$

If $\beta > 1$, scaling is super-linear. Given our three dependent variables, this would mean that the more papers a university produces, the more citations per paper it will receive (Van Raan, 2013), the higher the proportion of international co-publications and university-industry co-publications.

Adding the discussed covariates *AGE, METROPOP, CAPITAL, TECHNICAL, MEDICAL* and a set of *COUNTRY* dummies to this equation results in the following testable model:

$$\ln Y_i = \ln\alpha + \beta_1 \ln SIZE_i + \beta_2 \ln AGE_i + \beta_3 \ln METROPOP_i + \beta_4 CAPITAL_i + \beta_5 TECHNICAL_i + \beta_6 MEDICAL_i + \sum_{j=1}^{26} \gamma_j COUNTRY_{ij} + \varepsilon_i \tag{3}$$

We apply OLS regression to estimate this model, given the normally distributed log-transformed dependent variable. Throughout the analyses, as assumed, the residuals were normally distributed.

[3] Yet, this may not fully control for differences in disciplines as shown by Bornmann et al. (2013). Using latent class analysis to categorize universities and research institutions into scientific areas, they showed that certain institutions have an advantage in the ranking positions when compared with others if specialized in certain subject areas. This advantage manifests itself also when a performance indicator is field-normalized.

Table 1
Descriptive statistics.

	Min	Max	Mean	Std. Dev.
ln(P_top10) - *CITATION*	4.443	9.551	6.386	0.918
ln (P_int_collab) - *INTERNATIONAL*	5.298	10.180	7.632	0.852
ln(P_UI_collab) - *INDUSTRY*	1.099	8.298	5.490	1.033
ln(size) - *number of publications*	7.440	11.010	8.610	0.680
ln(age) – *university age*	1.386	6.947	4.698	0.859
ln(metropop) – *metropolitan population*	8.424	17.390	14.220	1.828
Capital city	0	1	0.181	
Technical university	0	1	0.123	
Medical university	0	1	0.049	
Australia	0	1	0.031	
Austria	0	1	0.012	
Belgium	0	1	0.009	
Brazil	0	1	0.017	
Canada	0	1	0.036	
China	0	1	0.120	
Finland	0	1	0.009	
France	0	1	0.031	
Germany	0	1	0.063	
Greece	0	1	0.008	
India	0	1	0.023	
Iran	0	1	0.017	
Israel	0	1	0.009	
Italy	0	1	0.044	
Japan	0	1	0.049	
Netherlands	0	1	0.017	
Poland	0	1	0.009	
Portugal	0	1	0.008	
South Korea	0	1	0.040	
Spain	0	1	0.037	
Sweden	0	1	0.013	
Switzerland	0	1	0.009	
Taiwan	0	1	0.020	
Turkey	0	1	0.015	
United Kingdom	0	1	0.057	
United States	0	1	0.211	

As a robustness check, we also present the results for size-adjusted dependent variables, where the numbers of highly-cited publications, international co-publications and university-industry co-publications are divided by the total number of a university's publications.

4. Results

Table 1 provides the variables' descriptive statistics, with the dummy variables indicating the shares. Capitals dominate the city landscape, accounting for 18% of all universities. In the dataset, 12% are technical universities and 5% medical schools. The US has by far the most universities in the Leiden Ranking (21%), with China coming second (12%).

Appendix A contains the correlation between the variables (excluding country dummies). It is clear that the dependent variables are highly correlated, which is not surprising considering they are all based on publication counts, and thus highly size-dependent. Looking at the correlation between the independent variables, we observe only low correlations, indicating the absence of collinearity.

4.1. Univariate results

Starting with the univariate regressions following Eq. (2) with size as independent variable, we find evidence of super-linear scaling for all three indicators. The scaling coefficient for the number of highly-cited publications as dependent variable is $\beta = 1.279$ ($p < 0.0001$, $R^2 = 0.898$). This relationship indicates that if the size of a university increases by 1%, the number of highly-cited publications increases by 1.279%, which shows that larger universities excel more than smaller universities, on average. Scaling is also apparent when looking at international collaboration, although the effect is weaker, with a scaling coefficient of $\beta = 1.136$ ($p < 0.0001$, $R^2 = 0.823$). A 1% increase in university size would lead to 1.136% more international publications. Finally, looking at university-industry collaboration, we observe a strong scaling effect, with β=1.296 ($p < 0.0001$, $R^2 = 0.728$). This means a 1% increase in university size would lead to 1.296% more university-industry publications.

Table 2
Regression results.

	CITATION number of 10% highest cited publications		*INTERNATIONAL* number of international co-publications		*INDUSTRY* number of university-industry co-publications	
	Model 1a		Model 2a		Model 3a	
	Coef.	Std. Error	Coef.	Std. Error	Coef.	Std. Error
Constant	−3.950***	0.151	−1.094***	0.182	−4.762***	0.289
ln(*SIZE*)	1.288***	0.015	1.134***	0.018	1.304***	0.029
ln(*AGE*)	0.022*	0.012	0.029**	0.015	0.055**	0.023
ln(*METROPOP*)	−0.060***	0.006	−0.084***	0.007	−0.090***	0.012
CAPITAL	−0.057**	0.028	0.175***	0.034	0.089	0.054
TECHNICAL	0.070**	0.031	−0.019	0.037	0.213***	0.059
MEDICAL	0.078*	0.046	−0.185***	0.056	0.062	0.089
Country Dummies	No		No		No	
R^2	0.915		0.857		0.754	
	CITATION number of 10% highest cited publications		*INTERNATIONAL* number of international co-publications		*INDUSTRY* number of university-industry co-publications	
	Model 1b		Model 2b		Model 3b	
	Coef.	Std. Error	Coef.	Std. Error	Coef.	Std. Error
Constant	−3.661***	0.100	−1.208***	0.095	−4.512***	0.205
ln(*SIZE*)	1.182***	0.011	1.089***	0.010	1.142***	0.022
ln(*AGE*)	−0.017**	0.008	−0.006	0.008	0.020	0.017
ln(*METROPOP*)	−0.010**	0.005	−0.009*	0.004	−0.011	0.010
CAPITAL	−0.000	0.020	0.008	0.019	0.057	0.041
TECHNICAL	0.065***	0.020	−0.005	0.019	0.309**	0.041
MEDICAL	0.001	0.030	−0.074**	0.029	0.055	0.062
Country Dummies	Yes		Yes		Yes	
R^2	0.967		0.966		0.892	

*** significant at 1% level; ** significant at 5% level; * significant at 10% level.

4.2. Multivariate results

Table 2 shows the main results based on our multivariate regression analysis as specified in Eq. (3). Models 1*a*, 2*a* and 3*a* represent the results excluding country dummies and Models 1*b*, 2*b* and 3*b* results include country dummies. Note that in all the models, the R^2 is very high, mainly driven by the strong significance of the size variable and much less so by the inclusion of country dummies.

Model 1 gives the results regarding the number of universities' highly-cited publications, both excluding country dummies (Model 1*a*) and including country dummies (Model 1*b*). The size effect is pronounced in both models, but weaker once countries are controlled for. Age has a negative effect on a university's highly-cited publications once country dummies are included. This negative effect of age is surprising given the common opinion that more recently established universities cannot be expected to be on a par with much older universities. This suggests that the main difference between younger and older universities lies in their size, as older universities tend to be larger than younger ones (corr = 0.25). However, as size is controlled for, our results indicate that age in itself slightly hampers excellence. The negative effect size is however small indicating that a 1% increase in age, decreases the number of highly-cited publications only by 0.017%.

We obtained a similarly unexpected result regarding city size in Model 1*b*, as measured by the metropolitan population. Universities in larger cities do not perform better than universities in smaller cities. Once the size of the university is controlled for, city size in itself even slightly reduces a university's citation impact although the effect is very small. A 1% increase in population decreases the number of highly-cited publications by 0.010%. One reason for the small negative effect may be that universities in large cities experience diseconomies of scale, as universities in larger cities face more planning constraints and higher costs. Furthermore, students and staff may be reluctant to work in big cities because the living costs are higher, whereas wages are often determined nationally (Nomaler et al., 2014).

Concerning the disciplinary orientation of universities as captured by the *MEDICAL* and *TECHNICAL* dummies, we observe that technical universities have on average about EXP(0.065) = 1.07 times more highly-cited publications than generic universities holding all other variables constant. No significant effect is observed for medical universities. These results cannot be explained by disciplinary specializations as citation scores are field-normalized.

Turning to the results for international co-publications in *Model 2*, we observe again that size has a positive effect on performance while city size has a small negative effect on performance similar to *Model 1*. Medical universities appear to be less active in international collaboration compared to generic universities. On average these universities have EXP(-0.074) = 0.93 times less international co-publications compared to generic universities. This finding may well reflect the more tacit and applied nature of the knowledge they produce, as well as the stronger embeddedness of medical universities in national healthcare systems as compared to generic universities.

Finally, comparing the number of university-industry co-publications in *Model 3*, we see again a strong size effect. University age and city size do not affect the number of university-industry collaborations. We do however observe that university-industry collaborations are more common for technical universities (if country dummies are included). The effect is pronounced with technical universities having on average about EXP(0.309) = 1.36 times more university-industry publications compared to generic universities. These results are easily explained by the nature of the research conducted at technical universities (Csomos & Toth, 2016).

4.3. Country dummies

Looking at the country dummies as show in Fig. 1, taken from the regressions reported in Table 2 under Models 1*b*, 2*b* and 3*b*, we see considerable differences in performance across countries for all three performance indicators. Observing the confidence intervals shows that many coefficients for country dummies differ significantly from each other, as their intervals do not overlap. Bear in mind that we should interpret the coefficients by referring to every other country in the world not coded as a dummy, which means all countries with five or fewer universities among the 750 universities included.[4] By taking the exponent, we can ascertain the extent of the effect. For example, the value of 0.213 for Australia in Model 1*a* indicates that Australian universities have on average about EXP(0.213) = 1.24 times more highy-cited publications compared to the reference group, while the value of −0.404 for Brazil indicates that Brazilian universities receive on average 0.67 times the number of highly-cited publications compared to the reference group.

The upper panel of Fig. 1 shows the number of highly-cited publications, we observe the US and the UK as expected high-performers as well as the Netherlands and Switzerland. Interestingly, the UK and Switzerland outperform the US,[5] whereas the Netherlands' performance is similar to the US. This result is in line with country-level bibliometric analyses that show that the UK, Switzerland, and the Netherlands do well in terms of citation impact, and at least as well as (and usually better than) the US (Waltman & Van Eck, 2015). Our results, however, contradict the common emphasis in rankings for the top 10 best performing universities, where the US invariably dominates with eight or nine institutions (Li et al., 2011). Note also that France and Germany have intermediate scores. This should be interpreted against the backdrop of their respective national systems, where a relatively large proportion of top researchers work outside the universities in national research institutes. The low performers are all Asian countries, along with Poland and Brazil.

According to Li et al. (2011), a large amount of cross-country variation in university performance can be explained by just four socioeconomic factors: income, population size, R&D spending, and the national language. Indeed, with the exception of Japan and South Korea, citation performance seems to correlate with income levels, suggesting that differential funding levels can play a significant role. Our findings concur with previous results by Bornmann et al. (2013, 2014), who also found strong differences between universities across countries relating to– among other factors–differences in economic wealth (Bornmann et al., 2013).[6] The low performance in Japan and South Korea despite their economic wealth, calls for more in-depth research.

Model 2*b* and Fig. 1 (middle panel) on international co-publications reveals that smaller countries tend to be most active in international collaboration, especially multi-lingual countries such as Switzerland and Belgium, and to a lesser extent Austria. This effect is also visible from the observation that most coefficients have a negative sign suggesting that most countries perform less than the reference category mainly consisting of small countries, at least in terms of research output. Especially Asian countries are poorly connected internationally, which may be due to language barriers and their remoteness to the main centers of scientific research (US and Europe). The relatively low coefficient for the US may not come as a surprise because that country provides so many opportunities for domestic collaboration given its high number of research-active universities. Consequently, the amount of international collaboration is relatively lower than in smaller countries.

Model 3*b* and Fig. 1 (lower panel) on university-industry co-publications also shows large differences across countries. We see close relationships between universities and firms in North-Western Europe, but lower levels in Southern and Eastern Europe. As expected, we also see strong integration in Japan and South Korea, while emerging countries such as Brazil, China and India seem to have limited opportunities for university-industry collaboration. The differences between countries may not just point to institutional differences in providing incentives to collaborate with industry, but primarily to the existence

---

[4] Countries with five or fewer universities are: Argentina (3), Chile (2), Croatia (1), Czech Republic (3), Denmark (5), Egypt (3), Estonia (1), Hungary (5), Ireland (4), Malaysia (5), Mexico (2), New Zealand (5), Norway (4), Russia (2), Saudi Arabia (4), Serbia (1), Singapore (2), Slovakia (1), Slovenia (1), South Africa (5), Thailand (3), and Tunisia (1).

[5] To justify this claim we ran model1b including an additional dummy for the group of countries with five or fewer universities and taking US as the reference category. Both the coefficients for Switzerland and UK were positive and significant with respective beta-values (*p*-values) of 0.149 (0.066) and 0.059 (0.029), while the coefficient of the Netherlands was insignificant −0.011 (0.049).

[6] Bornmann et al. (2014) also showed large disciplinary differences in a wide range of subject areas. In some fields, national income is a clear determinant for research performance, while less so in other fields. For example, the highest publication rate for papers in medicine strongly depends on the GDP, whereas the best rate for mathematics papers is less strongly associated with GDP.

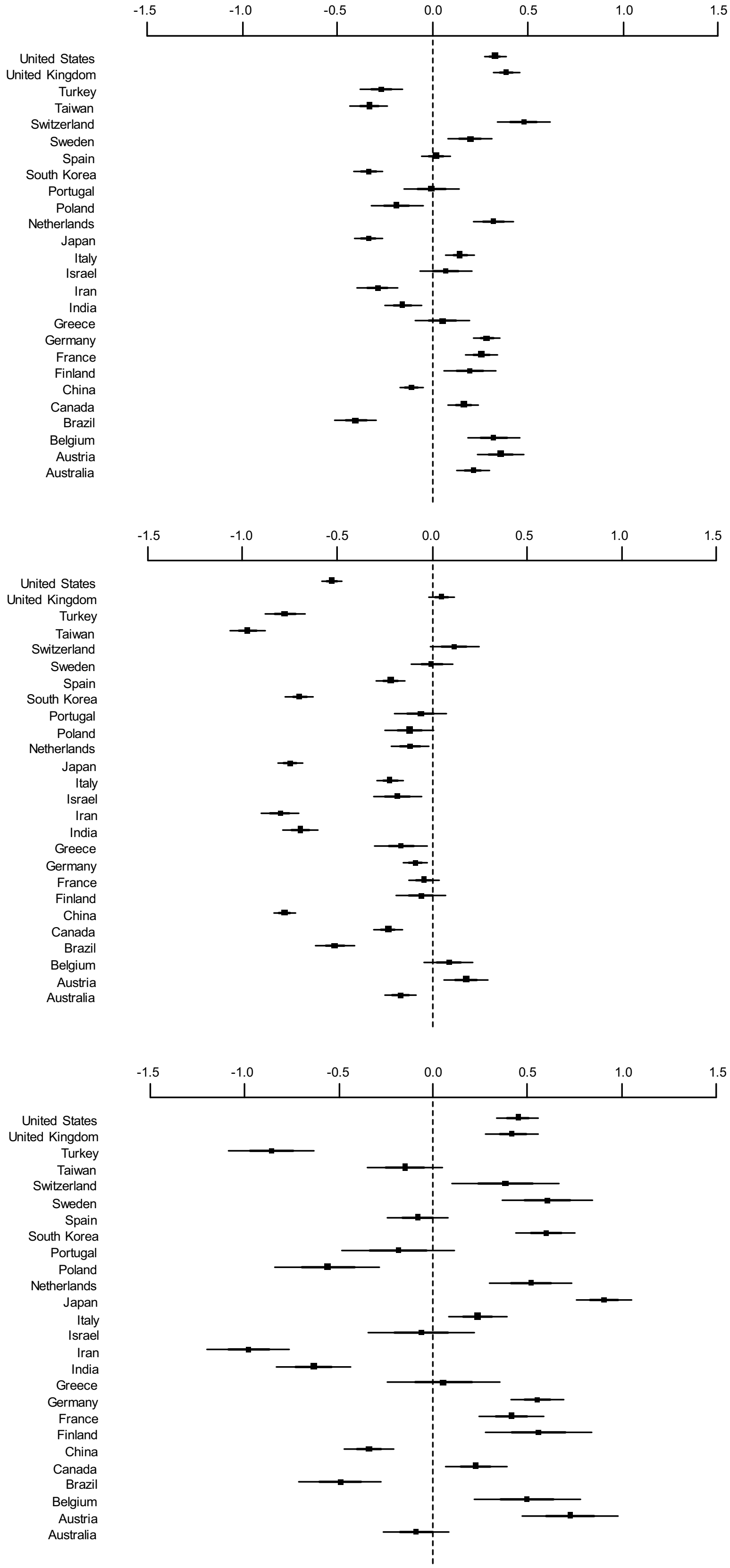

Fig. 1. Coefficient values of country dummies derived from Model 1b (upper panel), Model 2b (middle panel) and Model 3b (lower panel).

of local corporate R&D labs. Indeed, relatively small countries that nevertheless have a high performance in this respect are typically those with several science-based multinational firms (Finland, Netherlands, South Korea, Sweden, Switzerland).

In summary, we observe large country differences for each of the three performance dimensions. Yet, when comparing the results excluding and including country dummies, we see only a modest increase in the explained variance as expressed in the R^2. By contrast, also looking at the top ten percent highest cited publications per university, Bornmann et al. (2013) found that the variations between universities are mainly down to country differences rather than university differences. This discrepancy in results is probably due to the highly significant, *non-linear* specification of the size effect as expressed by the scaling coefficient. What is more, Bornmann et al. (2013) followed a different methodology by estimating multi-level logistic regressions on publications belonging to the top ten percent highest cited publications or not with university and country variables, while we estimate the total number of top ten percent highest cited publications per university and include country dummies.

Finally, the coefficients for the country dummies for each of the three dependent variables show very different patterns. Indeed, no single country excels in all three dimensions of citation impact, internationalization, and industry involvement. These results suggest that different national institutional environments favor specific strengths that may be partially mutually exclusive.[7]

4.4. Further analysis

Although the number of highly-cited publications captures a university's research excellence quite well, a simple alternative indicator would be to count the total number of citations a university receives. In Appendix B we repeated Model 1*b* analysis using the total number of citations instead of the number of highly-cited publications as a robustness check. Note again that the citations have been field-normalized. One interesting difference is that the scaling coefficient is smaller in the case of total number of citations (1.132) than for the number of highly-cited publications (1.183). This suggests that university size, as measured by the total number of publications, matters more for producing high-impact research than for generating impact in general. Regarding the other variables, the coefficients are all very similar in both regressions. This is not surprising given the strong correlation between the total number of citations that a university receives and the total number of publications that come under the ten percent most cited (corr = 0.995).

We also repeated the analyses presented in Models 1*b*, 2*b* and 3*b*, using size-independent variables as the dependent variables bydividing the original dependent variables by the total number of a university's publications (*SIZE*) and omitting the *SIZE* variable from right-hand side of Eq. (3). Appendix C presents the results. The size and significance levels of most variables remain unchanged compared to the original results presented in Table 2, but with lower R^2 values. The most robust are the size variable, the dummy for technical universities as well as the country dummies. The main differences lie in the impact of age and metropolitan population which sometimes lose their significance. The dummy for medical universities gained in significance in some models.

We finally looked at the interaction effect of size and age. These variables may be substitutes or complements. In so far as size and age signal reputation, a low value of one of these aspects may be compensated by a high value of the other, suggesting that the interaction effect is negative. On the other hand, size and age may also be complementary if a large size only signals reputation when combined with age. A negative and significant effect ($p < 0.01$) of the product ln(*SIZE*) and ln(*AGE*) was found when included in Model 3*b* with university-industry co-publications as dependent variable, also increasing the separate effects of size and age. The interaction effects were insignificant when included in Model *1b* and Model 2*b*.[8]

5. Conclusions

This empirical study analyzed university research performance in terms of research excellence, internationalization, and innovation. By means of regression analysis, we explained a university's research performance using three performance indicators from a set of structural variables including size, age, geography, disciplinary orientation, and country differences. In summary, the key robust results of our analysis of the factors driving university research performance are threefold: First, size matters for all three performance indicators. Large universities thus systematically over-perform in citation performance (indicating research excellence), in international co-publications (indicating internationalization), and university-industry co-publications (indicating innovation). Second, city size does impact performance - if any - in a negative way, as judged from the small negative association between city size and both highly-cited publications and international co-publications. The non-significant and even negative effects of city size question the notion that agglomeration economies exist in research (see Grossetti et al., 2014). Third, technical universities, which make up 12% of the universities worldwide, tend to outperform generic universities in terms of citation impact (unexpectedly) and university-industry relationships (expectedly). This underlines the specific character of technical universities compared to generic universities. We also do find that medical

[7] As noted previously regarding the US, international collaboration is relatively less relevant for large countries than smaller countries. Nevertheless, these indicators do serve as evaluation criteria for various university missions in most countries.

[8] The results can be obtained by contacting the authors.

universities underperform in terms of international co-publications which is perhaps expected given that a substantial part of medical research takes place within the context of national healthcare systems.

These performance indicators of excellence, internationalization and innovation are of course prone to conceptual ambiguity and uncertainty (Hicks, Wouters, Waltman, De Rijcke, & Rafols, 2015). Like all indicators, they only provide a limited and incomplete assessment of the phenomena of interest. Nevertheless, in many cases, these indicators are increasingly considered meaningful by students, university boards, and governmental bodies. Furthermore, the three university missions require very different policy interventions. Especially internationalization can be considered as network effects (Leydesdorff & Wagner, 2008), since there is no political institution mediating relationships.

Within this context, we indicate tentative explanations for the patterns found. In particular, we suspect that large universities benefit from economies of scale and greater visibility. This result underlines yet again that larger universities systematically over-perform in citation rankings. However, the exact cause remains under-researched. Though many may attribute the association between size and performance to economies of scale, especially in large-scale research projects and campus facilities, at least some of the effects may be due to sheer visibility as well as intra-university citation bias.

Regarding technical universities' differential performance. the higher number of university-industry relationships may well stem from the local, applied, and tacit nature of engineering research. More theorizing and in-depth analysis is required to further validate and scrutinize such explanations.

We also observed large differences between countries. By controlling for other variables, we identified high and low performing countries per indicator. At the same time, when looking at the three regressions jointly, we see that very few countries "excel" in all three dimensions. While these three dimensions feature prominently in many countries' university missions and policy documents, an interesting "trilemma" becomes apparent: no single country excels in all three dimensions of citation impact, internationalization, and industry involvement, with the possible exception of Switzerland. Notably, UK universities do particularly well in citation impact and internationalization, but are rather poor at industry involvement. Dutch universities, by contrast, do well in citation impact and industry involvement, but are not so internationalized as most other (European) countries. Sweden, in turn, does well in industry involvement and internationalization, but has a relatively low citation impact. These country patterns may point to trade-offs between performance criteria, with the nature of such trade-offs varying across countries. Clearly, the marked differences across countries call for more research on what type of macro-institutional and economic conditions affect university performance in terms of excellence, internationalization, and innovation (Bornmann et al., 2014).

From the perspective of countries, the results help us to establish the differences in university performance in countries with different institutional structures (Aghion, Dewatripont, Hoxby, Mas-Colell, & Sapir, 2008). The results suggest that various national institutional environments favor specific strengths related to scientific excellence, international collaboration or university-industry collaboration, which may be mutually exclusive. At the level of national research systems, the results also suggest that trade-offs exist between the universities' three potential objectives (excellence, internationalization, economic development). If so, countries need to make clear choices regarding the prioritization of citation impact, internationalization or industry involvement because it is unlikely that a national system can be designed institutionally that excels in all three dimensions.

One of the ways, then, to use our regression analysis to assess the performance of an individual university is to perform a residual analysis that compares the predicted value obtained from the regression coefficients with the observed value for that university (Bornmann et al., 2013). If the residual of an observation is positive (negative), it means the university in question is doing better (worse) than what we could expect from its structural features.[9] The results may also help us to use rankings for benchmarking purposes in a more informed way. From the perspective of universities, understanding the structural determinants underlying research performance allows them to select peers based on dimensions they deem relevant. For example, they can consider universities of a similar size as peers or, alternatively, compare with universities of a different size, while taking into account the non-linear scaling of performance with size. In accordance with Hicks et al. (2015), these factors could be supplemented with contextual information on a university's own strategy and mission, and with qualitative information regarding the institutional similarities with the university in question and its country location, compared to other universities it considers its peers. In this way, rankings can be a useful tool to benchmark a university against relevant peers, and to learn from well-performing universities in the peer group, as well as from the regional and national systems in which they operate.

Author contributions

Koen Frenken: Conceived and designed the analysis, performed the analysis, and wrote the paper.
Gaston Heimeriks: Conceived and designed the analysis, collected the data, and wrote the paper.
Jarno Hoekman: Conceived and designed the analysis and performed the analysis.

[9] An Appendix with the residual values of each universities from the estimated models 1b, 2b and 3b is available at: http://heimeriks.net/publications/.

Acknowledgments

The paper benefitted from comments during a seminar in the Science Policy Research Unit (SPRU) at the University of Sussex in February 2014, the EU-SPRI conference at Manchester Business School in June 2014, and the workshop on university rankings at CIRCLE, Lund University in September 2015. We would also like to thank Ben Martin and three anonymous reviewers for their comments, and Pieter van Os for his research assistance. The usual caveat applies.

Appendix A.

See Table A1.

Table A1
Correlation matrix.

	1	2	3	4	5	6	7	8	9
1. ln(P_top10) - *CITATION*	1.00	0.93	0.87	0.95	0.27	−0.07	−0.04	−0.09	−0.05
2. ln (P_int_collab) - *INTERNATIONAL*	0.93	1.00	0.82	0.91	0.29	−0.10	0.03	−0.11	−0.12
3. ln(P_UI_collab) - *INDUSTRY*	0.87	0.82	1.00	0.85	0.27	−0.08	0.00	−0.04	−0.06
4. ln(*SIZE*)	0.95	0.91	0.85	1.00	0.25	0.06	0.03	−0.10	−0.05
5. ln(*AGE*)	0.27	0.29	0.27	0.25	1.00	−0.15	0.02	−0.14	−0.10
6. ln(*METROPOP*)	−0.07	−0.10	−0.08	0.06	−0.15	1.00	0.42	0.15	0.14
7. *CAPITAL*	−0.04	0.03	0.00	0.03	0.02	0.42	1.00	0.13	0.05
8. *TECHNICAL*	−0.09	−0.11	−0.04	−0.10	−0.14	0.15	0.13	1.00	−0.09
9. *MEDICAL*	−0.05	−0.12	−0.06	−0.05	−0.10	0.14	0.05	−0.09	1.00

Appendix B.

See Table A2.

Table A2
Regression results for Total Number of Citations (TNCS).

	TNCS	no. of citations (ln)
	Coef.	Std. Error
Constant	−0.961***	0.070
ln(*SIZE*)	1.132***	0.007
ln(*AGE*)	−0.009	0.006
ln(*METROPOP*)	−0.006*	0.003
CAPITAL	−0.000	0.014
TECHNICAL	0.036**	0.014
MEDICAL	0.031	0.021
Australia	0.131***	0.030
Austria	0.201***	0.043
Belgium	0.208***	0.048
Brazil	−0.277***	0.037
Canada	0.087***	0.028
China	−0.126***	0.022
Finland	0.122**	0.048
France	0.154***	0.029
Germany	0.169***	0.024
Greece	0.031	0.051
India	−0.117***	0.034
Iran	−0.215***	0.037
Israel	0.027	0.048
Italy	0.088***	0.026
Japan	−0.250***	0.025
Netherlands	0.187***	0.037
Poland	−0.132***	0.048
Portugal	−0.035	0.051
South Korea	−0.245***	0.027
Spain	−0.002	0.027
Sweden	0.125***	0.041
Switzerland	0.336***	0.048
Taiwan	−0.249***	0.035
Turkey	−0.177***	0.039

Table A2 (*Continued*)

	TNCS	no. of citations (ln)
	Coef.	Std. Error
United Kingdom	0.265***	0.024
United States	0.217***	0.019
R^2	0.981	

*** significant at 1% level; ** significant at 5% level; * significant at 10% level.

Appendix C.

See Table A3.

Table A3

(a) Regression results for *CITATION*/P. (b) b Regression results for *INTERNATIONAL*/P. (c) Regression results for *INDUSTRY*/P.

(a)				
	CITATION/P	no. of 10% highest cited publications/no. of publications		
	Model A		Model B	
	Coef.	Std. Error	Coef.	Std. Error
Constant	−0.061***	0.012	0.075***	0.010
ln(*SIZE*)	0.022***	0.001	–	–
ln(*AGE*)	−0.002**	0.001	0.001	0.001
ln(*METROPOP*)	−0.001*	0.001	0.001*	0.001
CAPITAL	−0.000	0.002	0.001	0.003
TECHNICAL	0.006***	0.002	0.005*	0.003
MEDICAL	0.007*	0.003	0.003	0.004
Australia	0.019***	0.005	0.027***	0.006
	0.035***	0.007	0.032***	0.008
Belgium	0.034***	0.008	0.048***	0.009
Brazil	−0.037***	0.006	−0.035***	0.007
Canada	0.014***	0.005	0.023***	0.005
China	−0.013***	0.004	−0.016***	0.004
Finland	0.017**	0.008	0.024**	0.009
France	0.026***	0.005	0.033***	0.006
Germany	0.029***	0.004	0.037***	0.005
Greece	0.002	0.008	0.004	0.010
India	−0.014**	0.006	−0.025***	0.007
Iran	−0.024***	0.006	−0.033***	0.007
Israel	0.007	0.008	0.016*	0.009
Italy	0.011**	0.004	0.014***	0.005
Japan	−0.030***	0.004	−0.030***	0.005
Netherlands	0.036***	0.006	0.056***	0.007
Poland	−0.017**	0.008	−0.027***	0.009
Portugal	−0.006	0.008	0.003	0.010
South Korea	−0.031***	0.004	−0.030***	0.005
Spain	−0.001	0.004	−0.001	0.005
Sweden	0.017***	0.007	0.032***	0.008
Switzerland	0.061***	0.008	0.075***	0.009
Taiwan	−0.033***	0.006	−0.029***	0.007
Turkey	−0.022***	0.006	−0.032***	0.008
United Kingdom	0.045***	0.004	0.054***	0.005
United States	0.038***	0.003	0.050***	0.004
R^2	0.755		0.644	
Constant	0.364***	0.034	0.544***	0.027
ln(*SIZE*)	0.029***	0.004	–	–
ln(*AGE*)	−0.001	0.003	0.003	0.003
ln(*METROPOP*)	−0.004**	0.002	−0.001	0.002
CAPITAL	0.007	0.007	0.008	0.007
TECHNICAL	0.001	0.007	−0.001	0.007
MEDICAL	−0.018*	0.010	−0.023**	0.011
Australia	−0.083***	0.015	−0.072***	0.015
Austria	0.085***	0.021	0.082***	0.022
Belgium	0.059**	0.023	0.077***	0.024
Brazil	−0.219***	0.018	−0.217***	0.019
Canada	−0.112***	0.014	−0.101***	0.014
China	−0.290***	0.011	−0.293***	0.011

Table A3 (*Continued*)

(b)

	INTERNATIONAL/P	no. of international publications/no. of publications		
	Model A		Model B	
	Coef.	Std. Error	Coef.	Std. Error
Finland	−0.031	0.023	−0.022	0.024
France	−0.020	0.014	−0.010	0.015
Germany	−0.046***	0.012	−0.036***	0.012
Greece	−0.082***	0.025	−0.078***	0.026
India	−0.268***	0.016	−0.283***	0.017
Iran	−0.298***	0.018	−0.310***	0.019
Israel	−0.090***	0.023	−0.079***	0.024
Italy	−0.107***	0.013	−0.103***	0.013
Japan	−0.287***	0.012	−0.286***	0.013
Netherlands	−0.052***	0.018	−0.025	0.019
Poland	−0.070***	0.023	−0.084***	0.024
Portugal	−0.032	0.025	−0.020	0.026
South Korea	−0.272***	0.013	−0.271***	0.014
Spain	−0.110***	0.013	−0.110***	0.014
Sweden	0.004	0.020	0.024	0.021
Switzerland	0.079***	0.023	0.098***	0.024
Taiwan	−0.336***	0.017	−0.330***	0.018
Turkey	−0.280***	0.019	−0.293***	0.020
United Kingdom	0.033***	0.012	0.046***	0.012
United States	−0.224***	0.009	−0.208***	0.009
R^2	0.834		0.818	

(c)

	INDUSTRY/P	no. of university-industry co-publications/no. of publications		
	Model A		Model B	
	Coef.	Std. Error	Coef.	Std. Error
Constant	−0.001	0.009	0.031***	0.007
ln(*SIZE*)	0.005***	0.001	–	–
ln(*AGE*)	0.001	0.001	0.001**	0.001
ln(*METROPOP*)	−0.000	0.000	0.000	0.000
CAPITAL	0.001	0.002	0.001	0.002
TECHNICAL	0.015***	0.002	0.015***	0.002
MEDICAL	0.005**	0.003	0.005*	0.003
Australia	−0.010***	0.004	−0.008**	0.004
Austria	0.033***	0.005	0.032***	0.005
Belgium	0.019***	0.006	0.022***	0.006
Brazil	−0.020***	0.005	−0.019***	0.005
Canada	0.003	0.003	0.005	0.003
China	−0.017***	0.003	−0.017***	0.003
Finland	0.023***	0.006	0.025***	0.006
France	0.013***	0.004	0.015***	0.004
Germany	0.022***	0.003	0.024***	0.003
Greece	−0.005	0.006	−0.005	0.006
India	−0.022***	0.004	−0.024***	0.004
Iran	−0.030***	0.005	−0.032***	0.005
Israel	−0.009	0.006	−0.007	0.006
Italy	0.004	0.003	0.004	0.003
Japan	0.046***	0.003	0.046***	0.003
Netherlands	0.027***	0.005	0.032***	0.005
Poland	−0.021***	0.006	−0.024***	0.006
Portugal	−0.012*	0.006	−0.010	0.006
South Korea	0.024***	0.003	0.024***	0.003
Spain	−0.009**	0.003	−0.008**	0.003
Sweden	0.030***	0.005	0.033***	0.005
Switzerland	0.013**	0.006	0.016***	0.006
Taiwan	−0.011***	0.004	−0.010**	0.004
Turkey	−0.024***	0.005	−0.026***	0.005
United Kingdom	0.014***	0.003	0.016***	0.003
United States	0.016***	0.002	0.018***	0.002
R^2	0.646		0.639	

*** significant at 1% level; ** significant at 5% level; * significant at 10% level.

References

Aghion, P., Dewatripont, M., Hoxby, C., Mas-Colell, A., & Sapir, A. (2008). . *Higher aspirations: An agenda for reforming European universities* (Vol. V) Bruegel Blueprint Series.

Bonaccorsi, A. (2016). Addressing the disenchantment: universities and regional development in peripheral regions. *Journal of Economic Policy Reform*, (August 27), 1–28. http://dx.doi.org/10.1080/17487870.2016.1212711

Bornmann, L., Mutz, R., & Daniel, H. D. (2013). Multilevel-statistical reformulation of citation-based university rankings: The Leiden ranking 2011/2012. *Journal of the American Society for Information Science and Technology*, *64*(8), 1649–1658.

Bornmann, L., Stefaner, M., de Moya Anegón, F., & Mutz, R. (2014). What is the effect of country-specific characteristics on the research performance of scientific institutions? Using multi-level statistical models to rank and map universities and research-focused institutions worldwide. *Journal of Informetrics*, *8*(3), 581–593.

Csomos, G., & Toth, G. (2016). Exploring the position of cities in global corporate research and development: A bibliometric analysis by two different geographical approaches. *Journal of Informetrics*, *10*(2), 516–532.

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

Daraio, C., Bonaccorsi, A., & Simar, L. (2015). Rankings and university performance: A conditional multidimensional approach. *European Journal of Operational Research*, *244*(3), 918–930.

Etzkowitz, H., Webster, A., Gebhardt, C., & Terra, B. (2000). The future of the university and the university of the future: Evolution of ivory tower to entrepreneurial paradigm. *Research Policy*, *29*(2), 313–330.

European Journal of Education. (2014). Special issue: global university rankings. A critical assessment. *European Journal of Education*, *49*(1), 1–158.

Grossetti, M., Eckert, D., Gingras, Y., Jégou, L., Larivière, V., & Milard, B. (2014). Cities and the geographical deconcentration of scientific activity: A multilevel analysis of publications (1987–2007). *Urban Studies*, *51*(10), 2219–2234.

Heimeriks, G. (2013). Interdisciplinarity in biotechnology, genomics and nanotechnology. *Science and Public Policy*, *40*(1), 97–112.

Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). The Leiden manifesto for research metrics. *Nature*, *520*(7548), 429.

Kehm, B., & Stensaker, B. (Eds.). (2009). *University rankings, diversity, and the new landscape of higher education*. Rotterdam: Sense Publishers.

Leydesdorff, L., & Wagner, C. S. (2008). International collaboration in science and the formation of a core group. *Journal of Informetrics*, *2*(4), 317–325.

Li, M., Shankar, S., & Tank, K. (2011). Why does the USA dominate university league tables? *Studies in Higher Education*, *36*, 923–937.

Moed, H. F. (2017). A critical comparative analysis of five world university rankings. *Scientometrics*, *110*(2), 967–990.

Nomaler, Ö., Frenken, K., & Heimeriks, G. (2014). On scaling of scientific knowledge production in U.S. metropolitan areas. *PUBLIC LIBRARY OF SCIENCE*, *9*(10), e110805. http://dx.doi.org/10.1371/journal.pone.0110805

Olcay, A. O., & Bulu, M. (2016). Is measuring the knowledge creation of universities possible?: A review of university rankings. *Technological Forecasting and Social Change*, http://dx.doi.org/10.1016/j.techfore.2016.03.029, in press

Piro, F. N., & Sivertsen, G. (2016). How can differences in international university rankings be explained? *Scientometrics*, *109*(3), 2263–2278.

Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., & Stirling, A. (2012). How journal rankings can suppress interdisciplinary research: A comparison between innovation studies and business & management. *Research Policy*, *41*(7), 1262–1282.

Safon, V. (2013). What do global university rankings really measure? The search for the x factor and the x entity. *Scientometrics*, *97*(2), 223–244.

Shin, J. C., & Toutkoushian, R. K. (2011). The past, present, and future of university rankings. In J. C. Shin, R. K. Toutkoushian, & U. Teichler (Eds.), *University rankings. theoretical basis, methodology and impacts on global higher education* (pp. 1–16). Berlin: Springer.

Shin, J. C., Toutkoushian, R. K., & Teichler, U. (Eds.). (2011). *University rankings. theoretical basis, methodology and impacts on global higher education*. Berlin: Springer.

Van Raan, A. F., Van Leeuwen, T. N., & Visser, M. S. (2011). Severe language effect in university rankings: Particularly Germany and France are wronged in citation-based rankings. *Scientometrics*, *88*(2), 495–498.

Van Raan, A. F. (2013). Universities scale like cities. *PUBLIC LIBRARY OF SCIENCE*, *8*(3), e59384. http://dx.doi.org/10.1371/journal.pone.0059384

Vorley, T., & Nelles, J. (2008). Institutional development of and beyond the Third Mission. *Higher Education Management and Policy*, *20*(3), 119–135.

Waltman, L., & Van Eck, N. J. (2015). Field-normalized citation impact indicators and the choice of an appropriate counting method. *Journal of Informetrics*, *9*(4), 872–894.

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E. C. M., Tijssen, R. J. W., Van Eck, N. J., et al. (2012). The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology*, *63*(12), 2419–2432.

Zitt, M., Ramanana-Rahary, S., & Bassecoulard, E. (2005). Relativity of citation performance and excellence measures: From cross-field to cross-scale effects of field-normalisation. *Scientometrics*, *63*(2), 373–401.