Journal of Informetrics 12 (2018) 365–384

Contents lists available at ScienceDirect

Journal of Informetrics

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

Funding, evaluation, and the performance of national research systems

Ulf Sandström[a,*], Peter Van den Besselaar[b]

[a] Dept. INDEK, KTH Royal Institute of Technology, SE-114 28, Stockholm, Sweden
[b] Department of Organization Sciences & Network Institute, VU University Amsterdam, Netherlands

ARTICLE INFO

Article history:
Received 11 August 2017
Received in revised form 15 January 2018
Accepted 15 January 2018
Available online 22 February 2018

Keywords:
Research policy
Input-output studies
Performance-based funding
Research efficiency
Bibliometrics
Citations

ABSTRACT

Understanding the quality of science systems requires international comparative studies, which are difficult because of the lack of comparable data especially about inputs in research. In this study, we deploy an approach based on change instead of on levels of inputs and outputs: an approach that to a large extent eliminates the problem of measurement differences between countries. We firstly show that there are large differences in efficiency between national science systems, defined as the increase in output (highly cited papers) per percentage increase in input (funding). We then discuss our findings using popular explanations of performance differences: differences in funding systems (performance related or not), differences in the level of competition, differences in the level of university autonomy, and differences in the level of academic freedom. Interestingly, the available data do not support these common explanations. What the data suggest is that efficient systems are characterized by a well-developed *ex post* evaluation system combined with considerably high institutional funding and relatively low university autonomy (meaning a high autonomy of professionals). On the other hand, the less efficient systems have a strong *ex ante* control, either through a high level of so-called competitive project funding, or through strong power of the university management. Another conclusion is that more and better data are needed.

1. Introduction

What are the characteristics of research systems that influence efficiency? That issue has a long tradition in the political economics of science: how to get most value for money, i.e. how to get the best possible results from the investments in research (Stephan, 2012). How these investments should be measured is a difficult issue, especially when the aim is international comparison (Luwel, 2004). Also, output of research is heterogeneous and one may take into account various dimensions of research activities (Bonaccorsi & Daraio, 2004, p.60). In this paper, we restrict the analysis to scholarly output, in terms of (field normalized) highly cited scientific publications, which are considered to represent the important contributions to the growth of knowledge. It should be noted that we do not discuss efficiency in producing e.g., societal relevant knowledge, or patents, or the number of papers in general. Apart from presenting some solution for *measuring the investments in science*, we will address the question *what factors determine the efficiency of research systems?*

* Corresponding author.
E-mail addresses: ulf.sandstrom@indek.kth.se (U. Sandström), p.a.a.vanden.besselaar@vu.nl (P. Van den Besselaar).

https://doi.org/10.1016/j.joi.2018.01.007
1751-1577/

Our paper is first and foremost a critical discussion of the *existing theories* and *dominant ideas* in terms of the *evidence currently available*, evidence that is often used in science policy discussions. As we shall show, the available data contradict several of the popular (theoretical) claims, and lead to other intelligible findings. However, given the problematic nature of the available data, we do not claim that we have robust findings, but we do have an interesting research agenda: Additional research is needed, especially there is a need for more prudence in data collection.

The contributions of this paper are: (i) we show that most of the popular claims about what makes an efficient research system should not be believed too easily, as they lack support, even from the data they are based on; (ii) we argue that several of the core concepts in theorizing efficient research systems are more problematic than realized in contemporary discussions; and (iii) we suggest what alternative approaches may provide better explanations of efficiency differences. Last but not least, we argue (iv) that more and better data are needed to investigate how structural characteristics of science systems influence efficiency.

2. Explaining efficiency

Several properties of science systems are associated with the idea of efficiency (Sandström & Heyman, 2015). (i) The structure of *research funding* has been emphasized, and especially the growth of *competitive project funding* at the expense of institutional funding is generally seen as a stimulus for efficiency. (ii) Also the introduction of *national research evaluation systems* is expected to increase performance and efficiency of science systems, as is (iii) the turn to *New Public Management* (NPM) with its performance contracts and performance based institutional funding (Auranen & Nieminen, 2010). The latter is often based on indicators and 'funding formula' (Jonkers & Zacharewicz, 2015) which may be based on a national evaluation system. (iv) NPM should improve efficiency, accountability (Schubert, 2009), and responsiveness to changes in the environment, requiring more *autonomous* universities with powerful managers. Finally, the literature in science policy studies often emphasizes (v) the role of *academic freedom* (Heinze, 2008). We will briefly discuss the underlying theories below. To summarize, the following factors explain efficiency differences between science systems:

(a) The level of competition
- Share of project funding
- Performance based funding systems
- National evaluation systems

(b) The level of university autonomy
- Financial, organizational, staffing and academic autonomy

(c) Academic freedom

2.1. The role of competition and evaluation systems

Competiveness is generally defined in terms of the share of *basic university funds* (i.e. General University Funds GUF, Institutional Funds or Block Grants) in total research funding. The higher the share of such institutional funding and consequently the lower the share of project funding, the less competition would exist in a research system (Abramo, Cicero, & D'Angelo, 2012). However, increasingly also institutional funding is based on performance whereas in the past it was mainly input based (e.g., student numbers). Also other system pressures, such as excellence initiatives, NPM, and national research assessments are associated with the level of competition (Auranen & Nieminen, 2010).

An interesting attempt to build a dataset for seven European countries plus Australia was done by Auranen and Nieminen (2010), without justifying the selection of countries. In their analysis they proposed a two-dimensional typology of (i) input versus output oriented institutional funding,[1] and (ii) the share of external (project) funding in total university funding. The UK is in their view an example of a highly competitive system as it combines a high level of project funding with output-oriented institutional funding through the REF/RAE, and the former by the large share of money that goes through the various UK research councils and charities. On the other hand, countries like the Netherlands and Sweden were classified as poor performers with low efficiency, i.e. high cost per paper, in "a quite non-competitive environment". The latter claim is based on the observation that institutional funding is input oriented (student numbers, history and politics) although in Sweden the level of project funding is considerably high (Van Steen, 2012). Finland, Australia and Denmark were positioned in an in-between group. Germany and Norway were a bit closer to Sweden and the Netherlands.

The question why competition would lead to higher performing systems is addressed by Abramo et al. (2012), who formulated a theory concerning the expected effects of scholarly competition on the structure and performance of the academic system: Over time competitive arrangements are expected to redistribute high performing scholars between universities, i.e. the competitive process should lead to a concentration of funding to the best scholars in a few top universities. They argue that this will lead to (i) a higher performance variety between universities and at the same time to (ii) a lower

[1] This distinction was first proposed by Jongbloed & Vossensteyn (2001): the distinction whether public subsidies are based "[. . .] on input elements (i.e. indicators that refer to the resources used and/or the activities carried out by the higher education institutions) or output elements (i.e. indicators that refer to the institution's performance in terms of teaching and research)." (p. 128).

performance variety within each university – and this would (iii) overall result in a high performing system. Low competition in a research system would lead to the opposite pattern: performance differences between universities will be small (as there is a lack of concentration of top talent) but the performance differences within universities will be large – leading to a low performing system. Finally, the position of a more competitive research system in the international context will improve. In competitive systems, some universities will be excellent and attract even more outstanding researchers and students worldwide and therefore move up in the international rankings. So, (iv) more competition within a national research system leads to a higher position of the best universities of that country in the international rankings (Abramo et al., 2012). They illustrated the approach for the Italian case: Italy is considered as a non-competitive system because of the low share of competitive (project) funding, resulting in a relatively low performance, and with the predicted pattern of talent distribution among and within universities. The question of course is a whether their theory also holds when other countries are included.

Some authors hold an opposite view, emphasizing the rather conservative effects of performance evaluation through peer review in project funding. Heinze (2008), in his review of eight articles concludes that peer review tends to avoid risky and rewarding research, as it works against speculative, unorthodox and multidisciplinary research proposals. In other words, competitive distribution of research funds (generally based on peer review procedures) may result into mainstream, more applied and relatively rigid research. This in turn leads towards lowering the quality and weakening the innovative capacity in research (see also Thomas & Nedeva, 2012).

2.2. The role of autonomy

Another approach to explain differences in scholarly performance comes from Aghion, Dewatripont, Hoxby, Mas-Colell, and Sapir (2007), comparing European and US universities. Main differences were found in (i) the level of (especially private) funding and (ii) in the degree of autonomy from government. They conclude that increasing autonomy from government is good for performance, as better performing countries (with higher ranked universities) also score higher on autonomy. The underlying idea is that autonomous institutions face less legal and political constraints, can react quicker to changing external demand, and more easily adapt the internal organization to meet those changing demands.

In a subsequent paper Aghion, Dewatripont, Hoxby, Mas-Colell, and Sapir (2010) argue that performance of universities may not only be related to autonomy, but also to competition. The one without the other may work negatively on performance. Universities are more productive when they have autonomy and, at the same time, are forced to compete for research resources. Productivity was operationalized in terms of the score in the Shanghai ranking, whereas autonomy indicators came from a survey among universities, with a response rate below 40 %. The study found that factors such as *high pay flexibility* and *independent budget responsibility* correlate strongly with rankings. Whether this is a causal relation, was investigated by the same authors using data on American universities at state level. Despite the entirely different institutional arrangements in the US, the same pattern was found as for the European universities – which suggest a causal effect. However, using rankings as performance measure is problematic, as reliability and validity of the underlying concepts are strongly questioned (Hazelkorn, 2011).

Somewhat more recent, the European University Association (EUA) collected data on university autonomy, not using a survey but through a network of national experts providing the data (Esterman, Nokkala, & Steinel, 2011). Overall, the studies show similar patterns even if the results deviate for some countries. As the EUA dataset covers more countries, we use these data in this paper.

Four dimensions of autonomy were distinguished: organizational, staffing, academic and financial autonomy. The increase of autonomy from the state is often accompanied with the implementation of NPM arrangements that provide university's top management the instruments to govern and manage the organization (Esterman et al., 2011; Salmi, 2009). The combination of autonomy and NPM reflects a break with the traditional model of *academic self-governance* of academic affairs and *large state influence* on the administrative and financial aspects of universities (Schubert, 2009). This also becomes clear if one looks more into detail of how autonomy is defined:

Organizational autonomy refers to the freedom of the institution to shape its organizational forms and its governance systems, not 'hindered' by legal regulations. It should be noted that this concept of autonomy seems at least partly in contradiction to academic freedom, as the legal regulation of higher education often constitutes the power base of university professors against university management. Accordingly, increased autonomy of the institution goes together with decreased autonomy of the core university staff. *Staffing autonomy* has the same double nature — it deregulates the hiring and rewarding of staff, but also lowers the legal protection against being made redundant. The latter protection is often seen as part of academic freedom: the freedom to do science without being dependent on university management. *Academic autonomy* is mainly related to the start and termination of programs, to admission procedures and to the freedom to decide on student fees. Here too, academic autonomy transfers decision-making not only from the state to the university management, but at the same moment from the faculty to higher university levels. *Financial autonomy* is related to the possibilities of owning and trading buildings, to borrowing money, and to move funds between years.

2.3. Academic freedom

The above definition of autonomy shows that institutional autonomy generally means the autonomy of the university from the state, and this leads to an increased power of the universities' (top)-management. Larger power of top management

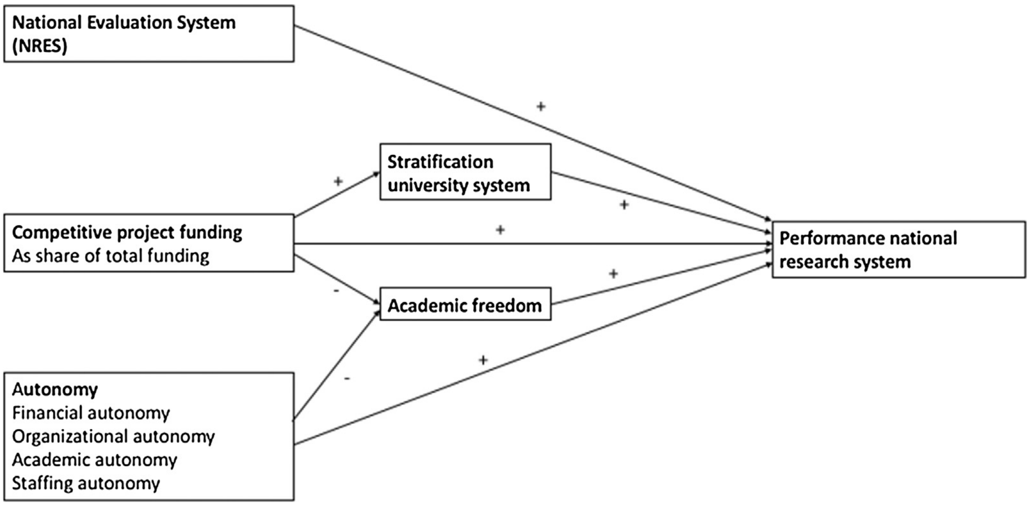

Fig. 1. The theoretical model based on the literature.

is often related to the introduction of forms of NPM leading to a decrease of the autonomy of the academic staff, which is also obvious in the EUA report. Several authors have argued that this may lead to less creative and more middle of the road and risk-avoiding research. This research will less often result in top cited papers, and therefore lowers the performance and the efficiency of the institution (Heinze, 2008; Luukkonen, 2012). Data on academic freedom are hardly available, but we will address this issue also in the empirical part, using the few data we could find.

Taking the mechanisms specified in these theories together leads to the following claims:

Hypothesis 1. Countries with a national research evaluation system (NRES) have a more efficient research system than those without a NRES.

Hypothesis 2. The higher the share of project funding, the higher the efficiency of the research system.

Hypothesis 3. The higher the universities' autonomy, the higher the efficiency of the research system.

Hypothesis 4. The higher the share of project funding, the more stratified the university system.

Hypothesis 5. The more stratified the university system, the higher the efficiency of the research system.

Hypothesis 6. The more academic freedom, the higher the efficiency of the research system.

Hypothesis 7. The higher the share of project funding, the lower academic freedom.

Hypothesis 8. The higher the universities' autonomy, the lower academic freedom.

These hypotheses are represented in the model in Fig. 1. The number of entities (=developed science countries) is too small to test the model statistically. Therefore, we will explore the (bivariate) relations between the variables, as explained in the next section, and with the data currently available. And even then, the statistics should be taken with care, given the data available.

3. Data and method

3.1. Input data

In order to account for efficiency, one needs reliable data about the input in the science system. Careful accounting of real R&D expenditures is needed, but obtaining reliable data has been a longstanding problem, which has dominated the political economics of research ever since the beginning of the discussions during the 1980s and 1990s (Cole & Phelan 1999; Luwel, 2004; Stephan, 2012). As countries differ in how they organize the science system, there are large structural differences in what counts as input. E.g., in systems where PhD students are employees and receive a salary, these salaries are an input; but if PhD students are treated as students, PhD students pay a fee and do not count as costs. Real estate is also treated differently between systems. In some countries, universities rent the premises, whereas in other countries they own the premises and this result in large differences in costs. In again other countries, the costs of buildings are carried by the state, and may remain invisible in the R&D inputs. Further differences are whether universities can shift funds between years or not, which also has impact on the way inputs are counted (Jacobsson & Rickne, 2004). Finally, academic salaries differ between countries, which make comparison of funding levels difficult. Using PPP does not help much as these are based on average price levels, and not on 'academic price levels' (Granberg & Jacobsson, 2006).

Table 1
Competitiveness of institutional funding.

COUNTRY	NRES[1]	Performance based institutional funding				Class[2]
	Started in	Hicks[6]	Auranen[3]	Stephan[4]	IPTS[5]	
Australia (AUS)	1993	yes	yes	yes	yes	yes-$
Austria (OS)						no
Belgium (BE)	1991	yes		yes	yes	yes-$
Canada (CA)						no
Denmark (DK)		[a]		[a]	[a]	no
Finland (FI)		[b]				no
France (FR)						no
Germany (GER)	2006			[f]		yes[f]
Ireland (IR)						no
Israel (IS)						no
Italy (IT)	2006	[c]		[c]	[c]	yes
Netherlands (NL)	1992					yes[g]
New Zealand (NZ)	2002	yes		yes		Yes-$
Norway (NO)	2006	yes	yes[d]	yes	yes	yes-$
Spain (SP)	1989	yes		yes	yes	yes-$
Sweden (SW)					[e]	no
Switzerland (CH)						no
United Kingdom (UK)	1986	yes	yes	yes	yes	Yes-$

[1] NRES: National Research Evaluation System. Only schemes introduced until 2006 are considered to affect publication strategies within the period until 2012. If no year is mentioned, no system is present.
[2] The three classes are: No: no NRES; Yes: NRES not related to funding; Yes-$ = NRES with funding relation.
[3] Auranen & Nieminen (2010).
[4] Franzoni et al. (2011).
[5] Hicks (2012).
[6] Jonkers and Zacharewicz (2015).
[a] In Denmark the system changed after 2008, too late to have effect in the period we study (de Boer et al., 2015).
[b] Hicks classifies the Finnish system as performance based. However, this is a misinterpretation when comparing it to available information in other sources (Auranen & Nieminen, 2010; Jonkers & Zacharewicz, 2015).
[c] Italy was classified as having performance based funding. However, the 2006 exercise had hardly effect on funding (Jonkers & Zacharewicz, 2015) and the performance based funding system was introduced only in 2009.
[d] Auranen & Nieminen classify Finland as non-performance based, but that reflects the situation before 2006.
[e] The Swedish performance based system was introduced only in 2009.
[f] The so-called *Excellence initiative*. It is an evaluation system, but as only a little bit of funding is related to it, we score it as a "1". For further information on the German situation, see Orr et al. (2007).
[g] see Moed et al. (1999), Van Steen and Eijffinger (1998), Westerheijden (1997).

Given this kind of differences between science systems, funding *levels* are difficult to compare. However, under the assumption that in the 2000–2009 period no structural changes took place within the set of countries in this study, the *change of inputs* can be compared. Efficiency can then be measures in terms of *change,* instead of *levels*.

Our approach focuses on change, and consequently the structural differences between countries are less important. Overall, if the rate of change of the variables is used instead of the levels, much of the effects of accounting and statistical differences between countries will be eliminated, and only changes *within* each country during the time period studied may affect the comparison.

National systems were, from an institutional point of view, rather stable in the period studied – after considerable changes during the 1990s (Senker, 1999). We do not take a wide variety of policy changes into account here, following the hypothesis of Bonaccorsi (2007) that institutions and institutional change are dominant. Where institutional changes did take place, data were adapted for that (see Appendix A). Several sources were used to find out whether structural changes took place (e.g., Reale, 2017). From country studies we know whether the institutional make up has changed, which was the case in Denmark where the universities merged and where public research institutes were integrated in the HE sector. In some countries, a NRES was introduced, such as in Norway. In other countries, the NRES was introduced too late to have an effect on the period studied here (see Table 1). The PREF study (Public Research Funding) shows that in most countries (where data are available), the competitive share of institutional funding was stable (Reale, 2017). Data on changes in the share of project funding between 2000 and 2009 are unfortunately not available. More details concerning data can be found in the Appendix A.

As we use the rate of change in input and output, it is important to use a fairly long period of time. Our dataset covers the period 2000–2009 and consist of 32 countries, for which economic data where present and publication data have a reasonable magnitude. We use of these 18 developed countries (see Appendix A for explaining the selection). Publication and citation data are used of the period 2002–2011 as we apply a two-year time lag between funding and output (c.f. Crespi & Geuna, 2008).

One may assume that countries that start at a low efficiency level may easier improve over time, as these systems have more suboptimal structural characteristics. However, that requires reorganizing in order to increase efficiency – which is not easy and not often done. So, in fact we have a set of structurally different countries, which can be compared in terms of efficiency. This can be done under the assumption of absence of economies of scale: we assume that the effect of an *increase* of funding is *not related* to the level of funding. Then we can compare the levels of input change between countries, without taking the level of funding into account (which is important, as good data for comparison are anyhow not available). So, in a more efficient research system, growth of funding leads to a stronger increase of output than in a less efficient system. Given this, we calculate the expected output increase using regression. The difference between expected and realized output increase is a measure for the efficiency of the system.

Interesting contributions to the discussion stem from Pan, Kaski, and Fortunato (2012) and from Cimini, Zaccaria, and Gabrielli (2016). The former conclude that output increases linear with input, and that there is an investment threshold in order to have an above world average impact. These findings seem plausible, especially as they include a heterogeneous set of countries in their analysis with research systems at very different levels of maturity. Although we would argue against using absolute levels of HERD (or GOVERD, BERD) in this type of analysis, as these data are very difficult to compare, the large differences between less and more developed science systems may reduce the effect of the differences in the data definitions and collection methods between countries.

The latter introduces another important dimension, which is the disciplinary structure of the national research portfolio. Indeed, as types of research may have different cost structures, the change of the research portfolio towards 'cheaper research fields' may influence measured efficiency (Abramo and D'Angelo, 2017; Aksnes, Sivertsen, van Leeuwen, & Wendt, 2016). And, developed research systems are much more diversified than those of the developing research systems, and the underdeveloped systems. Elsewhere, it has been shown that (i) diversification increases related to the size of the systems, (ii) that several typical specialization models exist in parallel to each other, but (iii) that the developed countries show a rather similar field composition and relative specialization (Horlings & Van den Besselaar, 2013). Finally, the price differences may be more related to the fixed costs than to the marginal costs, e.g. when a country enters more massively into a new field. In order to reduce the effect of differences in specialization patterns, only developed countries were included in this study.

3.2. Output data

How to measure output is another fundamental question: Auranen and Nieminen (2010) proposed the use of publication counts and citations counts as proxy for volume and quality of output (see also Cimini et al., 2016; Pan et al., 2012). We suggest a different approach, as we only take into account the 10% most cited papers (per subject category and document category), as these are the real contributions to scientific progress (Moed, van Leeuwen, & Visser, 1999; Tijssen, Visser, & van Leeuwen, 2002; Waltman et al., 2012), whereas neglecting the rest.

Data was collected from the *Web of Science* web interface, and further processing was done by the authors. We use the fractional count of field normalized 10% most cited papers to measure scientific output. This provides us with a *size-dependent* indicator of performance, an indicator sensitive to the fact that more contributions also indicate a higher performance. This was done to avoid an overestimation of the output of small countries that collaborate with large countries. In the remainder of the paper, we refer to this performance indicator as *FP10%*.

Furthermore, it is preferable to use non-parametric indicators – like the FP10% – as these are not sensitive to the skewness of citation distributions. The indicators are calculated using citations of the following document categories: Articles, Letters, Proceeding Papers and Reviews, covering the journals indexed in the databases SCI-Expanded, SSCI and A&HCI. Self-citations have been taken away based on first author name. When someone is first author of an article (A) and later also on a subsequent article (B) that refers to the former article (A), then it is considered as a self-citation. The citation window covers the year of publication up until Aug 1, 2014. Furthermore, in order to cover only international journals, we have applied the idea proposed by CWTS to exclude *non-core* journals.[2] Only journals that meet the following criteria are included: 1) publish in English, 2) have an international scope as reflected by the countries from which authors come and from where citations are received; 3) should have references to other core journals. This takes away a number of the local and regional journals that have been accepted by WoS during the more recent years. This has mainly an effect on measuring performance of countries that only recently have entered the science system – and these are not in our sample. It should be noted that the Leiden list of core journals applies to the situation in 2012–2013 and that we use this list for historical data. For the calculation of FP10% articles we apply the method proposed by Waltman and Schreiber (2013).

3.3. Competition

The *level of competition* in the science system is measured in different ways, as competition may take place between organizations and between individuals, and above that, several competitive mechanisms may coexist. Competitiveness relates to the type of available research funds and to the regulations for these funds. There is considerable confusion over categories in the OECD statistics: competitiveness concerns the research money that comes as variable resources, from competitive schemes organized by *second stream organizations* like research councils, foundations or government agencies.

Firstly, we measure competition in terms of competitive project funding as share of total funding. The higher the share of competitive funding and the lower the share of institutional funding, the higher the level of competitiveness is in a country.

[2] http://www.leidenranking.com/methodology/indicators.

In order to measure the share of institutional funding and project funding, we do not use the OECD GUF figures[3] but the data collected by Van Steen (2012) who specifically investigated the share of project funding in several countries, using a fairly consistent methodology. We assume that these data better measure the structure of national R&D funding.[4] The report by Van Steen (2012) does not include all OECD countries, so we had to add data for some countries using GUF figures (OECD, 2012), and (in case of Sweden) national statistics.

Secondly, we distinguish between countries that have a national research assessment system and those countries that have not, as the existence of such systems may make institutional funding more competitive too. If such a system exists, it may (e.g. UK) or may not (e.g. Netherlands) relate directly to funding.

Finally, we classify the countries in terms of whether institutional funding is based on research performance evaluation, as defined by Hicks (2012). There are several sources that take sometimes similar and sometimes slightly different approaches to defining 'performance based' (e.g., Auranen & Nieminen, 2010; Dawson, van Steen & van der Meulen, 2009; Franzoni, Scellato, & Stephan, 2011; Geuna & Martin 2003; Jonkers & Zacharewicz, 2015), but comparing the assessments in several of those sources leads to similar classifications as Table 1 shows. Please note that these sources also refer to each other. We distinguish three groups (last column of Table 1): Countries without a NRES, countries with a NRES with is not related to funding, and a group countries where institutional funding is related to a NRES. Within the last group, it may make a difference what part of institutional funds is competitive. Available data do not cover all countries and do not always cover our period. Hence, the comparability remains unclear (Reale, 2017). Nevertheless, the share of competitive institutional funding seems unrelated to the dependent variable (efficiency) in this study.

3.4. Autonomy

We use data on autonomy from Esterman et al. (2011) covering 28 European countries. For each of the countries, specialists were asked to complete the questionnaire that consisted of 24 items covering four dimensions of autonomy of the universities in relation to government. *Financial autonomy* was measured by six items such as: duration and type of public funding; ability to keep surplus; ability to borrow money; ability to own buildings; ability to decide on student fees. *Staffing autonomy* was measured through four items: capacity to decide on recruitment procedures, dismissals, promotion and salaries for senior staff and faculty. *Organizational autonomy* was measured through the following items: selection procedures, criteria, dismissal and terms of office of the executive head; inclusion and selection of external members of governing bodies; capacity do decide on academic structures; capacity to create legal entities. *Academic autonomy* was measured through seven items as: capacity to decide on student numbers; to select students; to introduce programs (BA, MA, PhD); to select quality assessment mechanism; and to design the content of degree programs. The different autonomy dimensions do correlate moderately high, around 0.5, and are shown in Table 2. One should keep in mind that these are self-reporting data.

3.5. Academic freedom

We use data on academic freedom from a large study on changes in academic profession and working conditions (Teichler, Arimoto, & Cummings, 2013, p. 186). These data were collected in the middle of the period we study. Respondents were asked to express (dis)agreement (5 points Likert scale) on items like "I am kept informed about what is going on at this institution"; "Lack of faculty involvement is a real problem"; "Students should have a stronger voice in determining policy that affects them"; "The administration supports academic freedom"; and "Top-level administrators are providing competent leadership". The academic freedom data are in Table 2, last column.

3.6. Stratification of the university system

As specified in the theory section, we need indicators for the level of heterogeneity within the university system: How strong are the *quality differences between universities*, and how strong *within* universities. For calculating these we use the Leiden Ranking 2014, reflecting the situation at the end of the period under consideration. The Leiden Ranking gives the share of top 10 % highly cited papers for universities as a whole, and for the individual disciplines. Here, we calculate the coefficient of variance (CoV) at the country level.

For the *quality differences within universities*, we also use the Leiden Ranking, and we use of the score for top 10 % highly cited papers scores per broad discipline: (i) Cognitive sciences (ii) Earth and environmental sciences, (iii) Life sciences, (iv) Mathematics, computer science & engineering, (v) Medical sciences, (vi) Natural sciences, (vii) Social sciences. In qualitatively homogeneous universities, one expects that they have about the same ranking in all disciplines; in qualitatively

[3] In theory, it should be quite simple to distinguish between different funding streams going to university R&D, but that is not the case. Although OECD has been collecting data from national statistical offices for more than 40 years there is still remaining questions concerning the comparability of these data, not to mention the quality. For countries with a federal political structure the reporting of the General University Funds (GUF) does not seem to work, e.g. Belgium, New Zealand and Canada have the problem that GUF is measured as the sum coming from the federal government not including the fixed sum coming from local or regional government.

[4] Usually the university GUF is calculated based on the figure for General University Funds (part of *Civil GBAORD*) divided with the figure for HERD. These country statistics is a contested area as there are probably differences in how concepts behind the statistics are interpreted in each country.

Table 2
Basic data[a].

Country	Competitiveness		Autonomy[b]					Academic Freedom[c]
	NRES[d]	Share project funding[e]	Academic	Staffing	Organization	Finance	Average	
1 Australia	Yes-$	47.1						45
2 Austria	No	28.9	72	73	78	59	71	
3 Belgium	Yes-$	55.7						
4 Canada	No	34.6						61
5 Denmark	No	25.9	56	68	94	69	72	
6 Finland	No	52.8	90	92	93	56	83	19
7 France	No	34.0	37	43	59	45	46	
8 Germany	Yes	34.6	72	86	83	71	78	37
9 Ireland	No	52.2	100	82	81	66	82	
10 Israel	No	35.7						
11 Italy	Yes		57	49	56	70	58	
12 Netherlands	Yes	27.8	48	73	69	77	67	49
13 New Zealand	Yes-$	41.0						
14 Norway	Yes-$	41.5	97	67	78	48	73	31
15 Spain	Yes-$	38.0	57	48	55	55	54	
16 Sweden	No	54.1	66	95	55	56	68	
17 Switzerland	No	23.2	72	95	55	65	72	
18 UK	Yes-$	56.4	94	96	100	89	95	41

[a] Empty cell indicates that data are not available.
[b] Level of university autonomy based on EUA (2011). The authors do not provide reliability analysis results.
[c] Country opinions on academic freedom based on Teichler et al. (2013), p 186. We used the data reported in the book, as the underlying data were not available.
[d] From Table 1.
[e] Share of project funding in total funding, based on Van Steen (2012). The figures reflect the 2008 situation. For countries missing in Van Steen (2012), we use adapted OECD figures on General University Funds (see Appendix A): France, Spain, UK. The figure for New Zealand is based on national sources. This counts too for Sweden, where the university hospitals were added.

Table 3
Output change by input change.

	Standardized Beta	t	Sig.	95% Confidence Interval for B	
				Lower Bound	Upper Bound
Model 1[a]					
(Constant)		0.252	0.804	−1.505	1.914
Increase funding 2000–2009	0.836	6.292	0	0.681	1.368
Model 2[b]					
(Constant)		0.416	0.683	−2.866	4.266
Increase funding 2000–2009	0.844	6.097	0	0.675	1.394
Efficiency level 2002[c]	−0.047	−0.339	0.739	−6.768	4.902

[*] Dependent Variable: Increase output (top cited papers) 2001–2011.
[a] Adjusted $R2 = 0.682$.
[b] Adjusted $R2 = 0.664$.
[c] Total number of top-cited papers (2001) in a country, divided by HERD, PPP normalized.

heterogeneous universities, one expects that the disciplines differ more in their rankings. We calculate the CoV for the top 10 % cited papers in the seven fields. The obtained university scores are then averaged to a score for each country.

Finally, we use as *top university* the countries' highest ranked university according to the Leiden ranking.

4. Findings

4.1. Measuring efficiency

Intuitively, the concept of efficiency is clear: the more efficient a system is, the more output is generated from a unit input. If comparable input and output data would exist, one could easily use the ratio as measure of efficiency. As this is not the case, we analyze the effect of the *change* of input on the *change* of output. Regression analysis shows that about 68 % of the variance of output change is explained by input change (model 1, Table 3). If we also include the efficiency *level* (the number of top cited papers per PPP euro at the start of the period under consideration), the explained variance remains the same, and the effect of the efficiency level is very small and not statistically significant (model 2, Table 3).

In other words, the starting level does not explain change: countries starting from low efficiency levels do not increase efficiency more easily. The *increase* of input explains about two third of the increased productivity, which is rather substantial, i.e. "money in – top cited publications out". Nevertheless, there remains quite some variance (within the set of developed countries 32 %) to be explained by other factors than funding, and these factors are the focus in the rest of this paper.

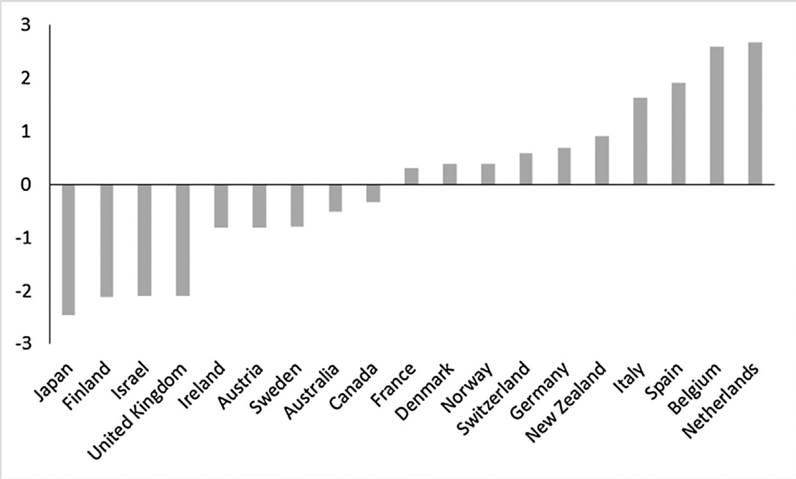

Fig. 2. Efficiency of research funding systems, funding period 2000–2009. Data: see Appendix A.

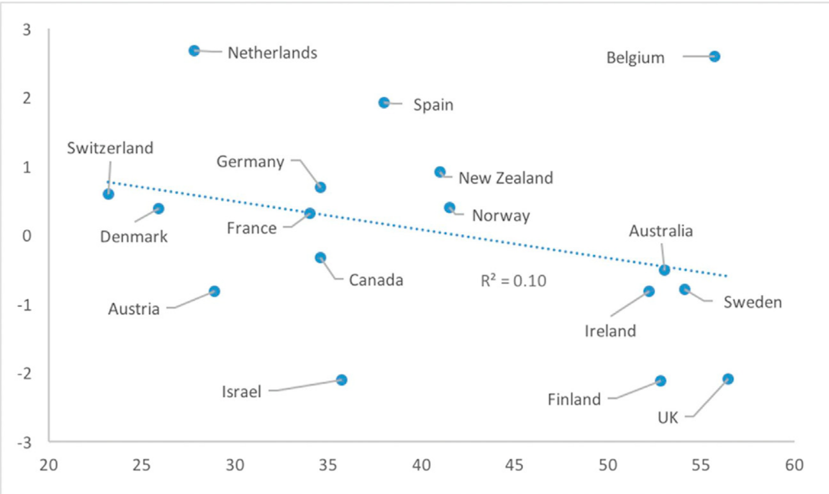

Fig. 3. Efficiency (Y-axis) by share of project funding (X-axis) 2000–2009.

Using the regression analysis of Table 3, we calculate the expected change in FP10% and compare that with the de facto change: Countries below the regression line are less efficient than countries above. We now operationalize our dependent variable 'efficiency' as the residuals (Fig. 2): the difference between de observed and the expected increase *FP10%*.

4.2. Efficiency and competition

It is often argued that the level of *competitive project funding* drives performance. Is this indeed the case? In fact, we find a moderate to small but negative correlation of about 0.3 between efficiency and the level of competitive project funding (Fig. 3). This finding seems to contradict many ideas about the positive effect any type of competitive funding would have on performance.

Distinguishing national systems in terms of the presence of a *national research evaluation system* (NRES) shows the following: (i) Countries with a (strong or weak) NRES score better in terms of efficiency than countries without such a system. This can be seen in Fig. 4 where the former countries (circles and rectangles) are in the higher part of the figure, and the latter (triangles) mostly in the lower part. The NRES systems score on average .91, and the non-NRES systems on average −.83 on efficiency; (ii) The negative correlation remains, and is actually stronger (−0.53 and −0.61) than for the whole sample (−0.3). And (iii) countries with a weak NRES (rectangles) score on average better on efficiency (1.61) than those with a strong NRES (circles) system (0.53). We support this visual evidence with a statistical analysis of the effect of the level of institutional funding on efficiency, using the existence (yes or no) of a national evaluation system as dummy: The share of institutional funding in total funding has a positive effect on efficiency of the system (Table 4). The same holds for the effect of an evaluation system – whether it is funding related or not.

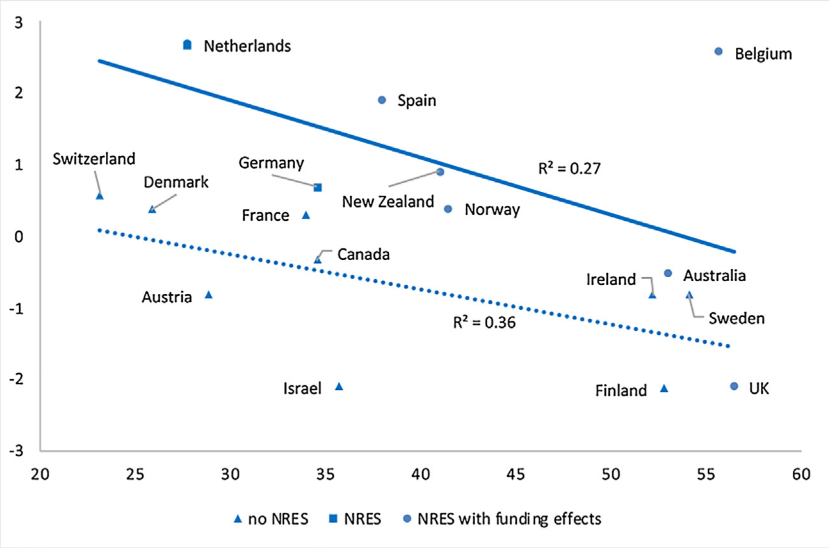

Fig. 4. Efficiency (Y-axis) byshare of project funding (X-axis) & by type NRES (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).
(red = no NRES; blue = NRES; green = NRES without funding effects).

Table 4
Performance (efficiency) by NRES[a] and by share institutional funding.

	Unstandardized Coefficients		Standardized Coefficients Beta	t	Sig.
	B	Std. Error			
(Constant)	−4.475	1.682		−2.661	0.019
National Research Evaluation System	1.800	0.582	0.624	3.090	0.008
Share institutional funding	0.062	0.026	0.474	2.345	0.034

[a] National Research Evaluation System; R2 = 0.389.

These results suggest that the NRES systems indeed function as an incentive to perform better, leading to a higher efficiency of the system. NRES systems may have created a strong performance culture leading towards better research, to more international publications and in turn this has geared the system towards higher (citation) impact (c.f. Sandström and Van den Besselaar, 2016). On the other hand, the positive correlation between the level of institutional funding and performance is interesting. It suggests that institutional funding is essential for a research system, and that increasing project funding may lead to over-competitiveness. To this conclusion there are exceptions requiring further study: e.g. Belgium which has a NRES system and a high level of project funding – but it scores very high on efficiency.

Concluding, Hypothesis 1 *is partly supported*, as evaluation systems have a positive effect on efficiency, but NRES without funding effects do better than funding related NRES; Hypothesis 2 *is not supported*, as more competitive project funding goes together with less efficiency.

4.3. Level of competitiveness and stratification of the university system

To investigate whether the stratification of the university system relates to competition and performance we distinguish the three aspects of stratification discussed above:

- The position of the highest ranked university
- The quality difference between universities in a country
- The quality differences within universities in a country

Table 5 shows the variables as explained in the methods section. Please note that the data for Italy are in line with Abramo et al. (2012) argument about Italy: differences between universities are comparatively low, differences within universities are comparatively high.

(i) Do highly competitive countries have higher ranked top universities? We investigate the relation between competitiveness and the performance of the highest scoring national university in the 2014 CWTS *Leiden ranking* – in terms of PP10%

Table 5
Quality differences between and within university by country.

Differences:	Highest ranked University[a]		Between university differences[b]	Within university differences[c]
Australia	13.3	ANU Canberra	0.160	0.285
Austria	12.2	KFU Graz	0.121	0.199
Belgium[d]	13.0	KU Leuven	0.129	0.210
Canada	13.8	U Toronto	0.186	0.219
Denmark	14.1	TU Denmark	0.121	0.260
Finland	10.5	U Helsinki	0.099	0.276
France	14.6	ENS Paris	0.167	0.222
Germany	14.2	TU München	0.141	0.235
Ireland	14.0	Trinity College	0.134	0.221
Israel	17.7	Weizmann Inst	0.419	0.255
Italy	11.6	U Trento	0.138	0.262
Netherlands	14.8	U Leiden	0.077	0.179
New Zealand	10.9	U Otago	0.181	0.249
Norway	9.9	U Oslo	0.030	0.234
Spain	11.8	Rovira i Virgili U	0.172	0.319
Sweden	12.7	Karolinska Inst	0.120	0.234
Switzerland	17.7	EPFL Lausanne	0.129	0.159
UK	18.4	U Cambridge	0.196	0.225

[a] Leiden Ranking 2014, PP10% score.
[b] Coefficient of Variation of PP10% scores of the universities within a country.
[c] Coefficient of Variation of PP10% scores for 7 domains of each university, and then averaged over universities in a country.
[d] Flemish universities only.

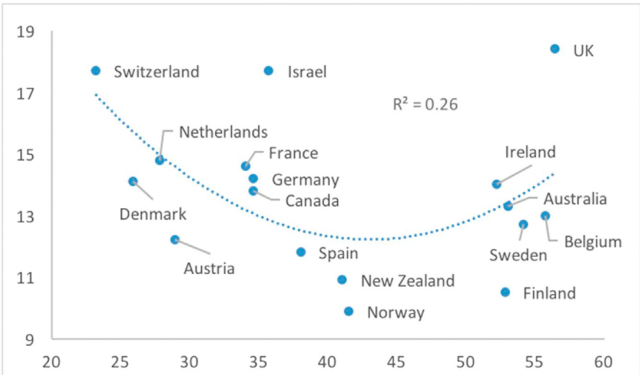

Fig. 5. Performance of the country's top* university (Y-axis) by share project funding (X-axis). *PP10%, Leiden Ranking 2014.

(time frame: 2009–2012). The top performers are universities in UK, Israel, and Switzerland, each with about a PP10% score of 18 %. The lowest is the best Norwegian university that had a PP10% score of 9.9 %.

The relation between the level of competitiveness (in terms on the share of project funding) and the performance of the countries' best university is represented by a u-curve (Fig. 5). High performing best universities are found within the UK with much competitive funding, but also in Switzerland with little competitive funding. Probably all these top-universities are not affected by competitive funding, as they can be expected to get 'competitive funding' on the basis of reputation.

- Does the level of competitiveness correlate with performance differences between universities? As explained in the data section, we calculated the coefficient of variation (CoV) of the universities' share in the top 10% cited papers. As Fig. 6 shows, the share of project funding does not correlate with differences between universities at the country level, and the explained variance is about zero.
- Does competitiveness correlate negatively with performance differences within universities, as Abramo et al. (2012) suggests? That is not the case: Fig. 7 suggests an inverted u-curve, with a long positive trend, and a modest negative trend at the end of the curve. We tested whether this is size-related (number of universities in the system), but that is not the case: the correlation is weak (0.192) and not significant ($p = .445$).
- Does the 'Abramo stratification' correlate positively with performance (in terms of efficiency) of the national systems? The correlation between the PP10% score of the *best university* and the efficiency score is negative (−0.26) and not positive as claimed by Abramo et al. (2012). The correlation of the *performance differences* between universities and efficiency is −0.40, whereas the theory of Abramo implies a positive relation. Only the correlation between the within-universities performance differences and efficiency is in line with Abramo's expectations, but also rather small (−0.13).

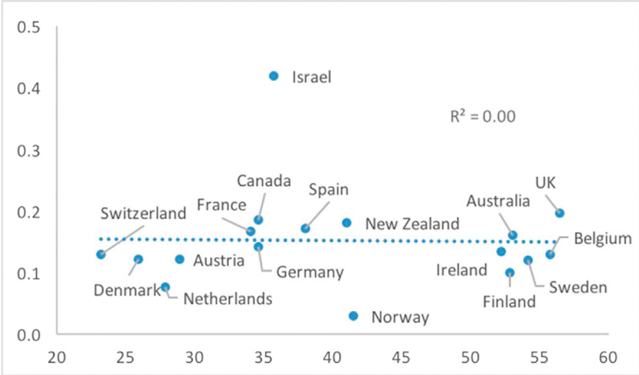

Fig. 6. Performance differences* *between* universities (Y-axis) by. Share project funding (X-axis). *CoV PP10% – Leiden Ranking 2014.

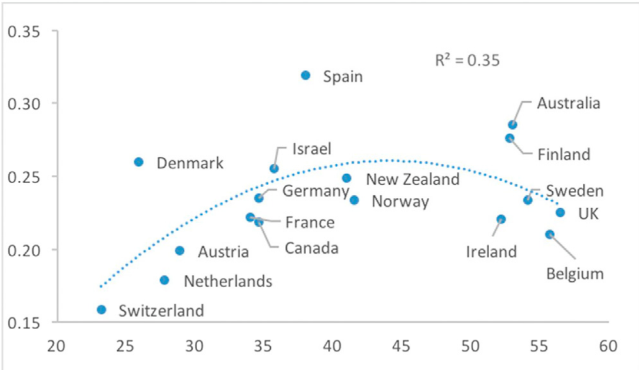

Fig. 7. Performance differences *within* universities* (Y-axis) by.
Share project funding (X-axis).
*Average of universities' CoV of PP10% – Leiden Ranking 2014.

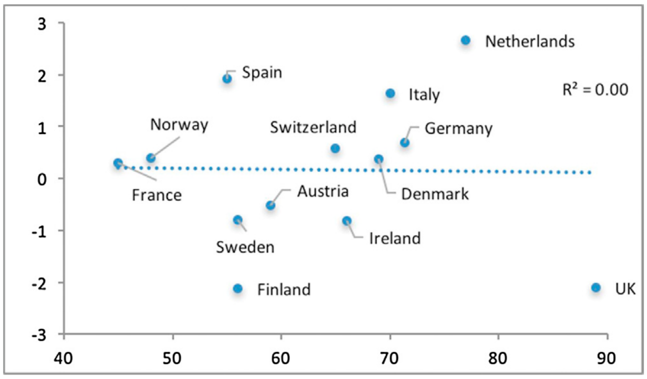

Fig. 8. Efficiency by financial autonomy.

Summarizing, the tests do *not support* Hypothesis 4: more competition does not result in more stratification. And stratification within the university system has a negative effect on the efficiency of the system: Hypothesis 5 in also not supported.

4.4. University autonomy and efficiency

An important possible explanation of differences between countries' efficiency of the research system is the level of university autonomy. We have data for four dimensions of autonomy, which are summarized in Figs. 8–11. Financial autonomy implies the possibility to spend money in different years may, which avoid institutions to spend funds for less relevant activities, simply because the money needs to be spent before the end of the year. However, contrary to expectations, financial autonomy seems to have no effect on efficiency (Fig. 8).

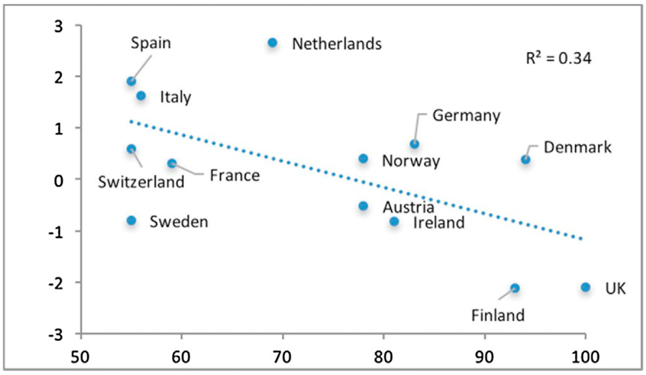

Fig. 9. Efficiency by academic autonomy.

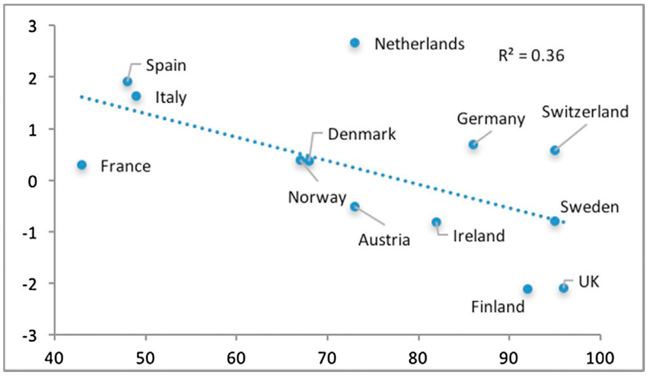

Fig. 10. Efficiency by staffing autonomy.

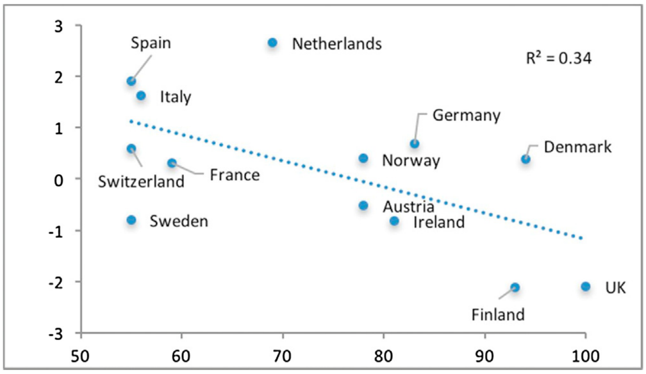

Fig. 11. Efficiency by organizational autonomy.

The other autonomy dimensions correlate negatively with efficiency. This holds for staffing autonomy (Fig. 9), for organizational autonomy (Fig. 10), and academic autonomy (Fig. 11). The higher the score on these dimensions of autonomy, the more that the university can decide on the structure of governance, the structure of the university and the working environments, and on teaching and student related issues, without interference of the government. However, there is a negative correlation with efficiency of the system.

As the sample is too small to include many variables, we did the statistical analysis for each of the autonomy dimensions: do the share of institutional funding, the existence of a NRES system, and one of the autonomy dimensions explain efficiency? For three of the autonomy dimensions there is no significant effect. Only in case of organizational autonomy we find a significant model: Organizational autonomy has a negative effect on efficiency; the existence of a NRES and the level of institutional funding have positive effects (Table 6). Overall, these data suggest that more independence for universities from the state does not lead to better performance. So, Hypothesis 3 *is not supported*, and in fact, the findings point in the opposite direction.

Table 6
Efficiency by evaluation system, institutional funding and autonomy.

	Unstandardized Coefficients Beta		Standardized Coefficients Beta	t	Sig.
	B	Std. Error			
(Constant)	−1.957	1.744		−1.122	0.295
Share institutional funding	0.069	0.018	0.576	3.735	0.006
NRES	1.376	0.415	0.491	3.313	0.011
Organizational autonomy	−0.037	0.014	−0.421	−2.713	0.027

Dependent Variable: Efficiency residuals. Adjusted R2 = 0.704.
NRES = National Research Evaluation System.

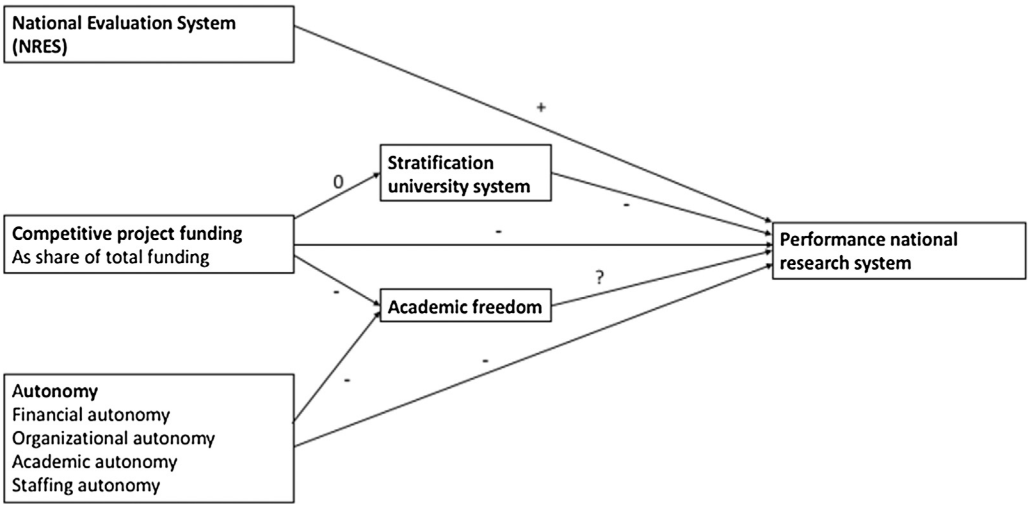

Fig. 12. The results of the analysis.

Autonomy of Higher Education organizations is defined in relation to government, but it also impacts within the HE organizations the relation between top management and the staff. The definition of autonomy clearly implies that more autonomy for the university goes hand in hand more managerial power at the expense of the role of the faculty. For example, systems where professors have a legally regulated position in university governance are considered less autonomous by the EUA study. In that sense, the findings about autonomy should be qualified: University autonomy seems to have a negative effect on performance, when combined with NPM instruments.

Based on the above reported findings, we now would expect a positive relation between academic freedom and efficiency – as is also suggested by several studies which are only based on interview data (Laudel, 2006; Heinze, 2008). Data at the national level about academic freedom hardly exist, apart from the survey based study of Teichler et al. (2013). This study covered quite a few countries, but data about academic freedom are only collected for twelve countries, of which not more than seven are also included in our study. Using these data, we find an almost zero correlation between academic freedom and efficiency, but as expected, a negative correlation between three of the four autonomy dimensions (financial autonomy as the only exception) and academic freedom, as well as a negative correlation between project funding and academic freedom. This supports Hypothesis 7 *and* Hypothesis 8, *but not* Hypothesis 6. However, the analysis of the effects of academic freedom suffer from the sparsely available data.

Fig. 12 summarizes the findings based *on the data available*. Given the low number of countries involved, these results should be used with extreme care.

5. Conclusions

Explaining efficiency needs adequate input and output data. Because input levels cannot be compared between countries, we focus on the relation between change in input and change in output: these variables are independent of structural differences between countries. National systems were rather stable in the period studied. With this approach we conclude that input changes in the science system determine to a large extent the changes in output. On the other hand, some 30 % of the variation in efficiency needs still to be explained by other factors. Based on the available data, we can shed some light on the explanations that are dominant in the literature. We do not claim to have solved the question of what factors do contribute to performance efficiency – but our analysis suggests that factors that are commonly believed to have a positive effect on efficiency in fact may not. This could either mean that competitiveness plays no role, or that competitiveness is not adequately measured by the shares of project and institutional funding. It also leads to new questions about the relation between the structure, stratification, and performance of research systems.

Our findings, based on available data, indicate that (i) Good functioning (efficient) science systems seem characterized by a well-developed *ex post* evaluation system combined with considerably *high institutional funding* and *not too large university autonomy* in the sense of managerial power. Whether autonomy from the state has a positive effect, when not combined with NPM but with *academic freedom,* is an important issue that remains unsettled because of the lack of data. Countries with an ex-post evaluation system are more efficient, but that is much less the case for 'strong' evaluation systems that have direct funding effects. (ii) On the other hand, the less efficient systems seem to have a strong *ex ante* control, either through a *high level of so-called competitive project funding*, and/or through strong power of *university management* in autonomous institutions.

What are the limitations? Further testing heavily depends on the availability of relevant data for more countries. (i) Although moving from *level* of funding to *change* of funding solves some of the data problem, for many countries there is still a need for data on the *share of project funding*, measured in a consistent way. The data developed by Van Steen (2012) cover only a subset of the relevant countries, and should be extended. It would also be better to have data for more moments in time. (ii) Competitive nature of institutional funding is only measured through the NRES. More details about the specific funding systems may be helpful, although difficult to obtain in a comparative mode. (iii) The output data as such are not problematic, as PP10% is a useful indicator for scientific output (Tijssen et al., 2002). However, it would be important to include also other types of output, as it is a highly relevant question how 'societal productivity' would depend on institutional and structural characteristics. It could be that optimizing different performance dimensions requires other (contradictory) system characteristics (Van der Weijden, Verbree, & Van den Besselaar, 2012). But comparable data for other output dimensions are not yet available. (iv) The data about the level of universities' autonomy are based on surveys and the quality of the survey as well as the response rates needs further attention. Even more importantly, one would prefer to use non-survey data about autonomy, e.g., based on a standardized interpretation of the formal and material relations between universities and governments in the different countries. For further research, one would also need data that measure autonomy at various moments in time, to better cover the relative long period under consideration. (v) The same holds for the data on academic freedom. (vi) As research in different fields may have different cost levels, the price per paper may differ between fields. Comparing efficiency may therefore need control for differences between national portfolios and changes in portfolios over time. This issue deserves further studies and new approaches (Abramo and D'Angelo, 2017; Aksnes et al., 2016). (vii) Finally, the coverage of countries is an issue – we could find data for the various variables, but not for all countries that could be included.

Apart from this, we need detailed case studies, comparing different countries. We are working on some cases: (i) The Netherlands, combining a moderate level of competitive funding with a NRES not related to funding, and a moderate level of autonomy, leading to an equal system with high efficiency. (ii) The UK, combining a high level of competitive funding, with a funding related NRES, and autonomous universities, resulting in a highly-stratified system with a low efficiency. (iii) Sweden, with relatively strong competition but no national evaluation system, resulting in a stratified reputation based system with a moderate efficiency. Doing these case studies may provide new perspectives on the nature and role of competition, autonomy and academic freedom. The next section gives some first thoughts.

6. Discussion

What is generally seen as higher levels of competition seem to go together with lower efficiency of the science system. Why would this be the case? Our findings point in two different directions: (i) Either the effect of competition has been misunderstood, or (ii) the concept of (merit based) competition has been incorrectly identified with specific funding and evaluation arrangements.

- A first reason why more competitive systems are less efficient may be related to the selection process. As many studies on peer and panel review have shown, ex ante selection of research projects does not function very well. For example, selection processes are not well equipped to select those with the best past performance (Bornmann, Stefaner, de Moya Anegon & Mutz, 2014; Hornbostel, Böhmer, Klingsporn, Neufield, & von Ins, 2009; Van den Besselaar & Leydesdorff, 2009) and predictive validity is even lower (for an overview: Van den Besselaar & Sandström, 2015). Panels obviously cannot distinguish between the good and the excellent. Underlying mechanisms are group processes in decision making (Van Arensbergen, van der Weijden, & van den Besselaar, 2014) which may result in conservatism (Hall, 1972; Heinze, 2008), nepotism (Sandström & Hällsten, 2008; Wennerås & Wold, 1997), sexism (Wennerås & Wold, 1997), and the unavoidable influence of reputation (Merton, 1968; Price, 1976).
- For similar reasons, autonomy of universities has a negative effect on performance. Autonomy of universities means that universities become organizational actors, with an increased power for management – at the expense of the influence of the academic staff. This results in ex ante steering, by managers that often belong to the same academic elite as those that populate panels (Musselin, 2013). And there is no reason why university managers would be able to select promising projects or research topics. Above this, as the discussions on managerialism versus professionalism have indicated, increased power of management at the expense of professional autonomy may lead to negatively impact the quality of research. NPM strategies tend to define research agendas top-down, or aim at defining the credible publication outlets

(the so-called *A-journal lists*). This is claimed to hinder interdisciplinary research, risky and innovative research (Nedeva, Boden, & Nugroho, 2012) and would result in middle of the road, mainstream, average quality research, based on restricted agenda's (Gläser & Laudel, 2016; Luukkonen, 2012).[5] Similar mechanisms as in the case of competitive funding may work here too, such as reputation based decision making, and distributional justice.
- If competitive funding and autonomy of universities have a negative relation with performance, does this mean that competition is irrelevant? We don't think so, but would argue that the nature of scholarly competition may be wrongly understood. As was already argued by Hagstrom (1964), competition is an essential part of science. He was the first to emphasize that discovery and achieving peer recognition are key drivers of research careers (Dasgupta & David, 1994; c.f. Stephan, 2012). Also, this is reflected by the traditional *credibility cycle* of the production, communication and collective evaluation of the results. More recently, the credibility cycle has been extended with the competition for research funds needed to carry out research (Garcia & Sanz-Menéndez, 2005; Latour and Woolgar, 1979). Our results actually suggest that the literature does not distinguish enough between *competition for resources* and *competition for results*. In fact, too much competition for resources (and especially for *funding*) seems on the macro level not to be beneficial for the performance of science system. This change in the meaning of competition is also visible in the changed meaning of acquiring funding: getting a prestigious grant is already seen as a performance, through its symbolic value – and what one does with it seems less important. Taking this together with the low (predictive) validity of grant decisions, a strong emphasis on competitive project funding may be a good recipe for low efficiency (c.f. Sandström & Wold, 2015).
- There is one more argument why competition and autonomy have negative effects on efficiency. The distinction between *prestige* and performance *excellence* (Paradeise & Thoenig, 2013), combined with the analyses of Whitley (2007), Musselin (2013), and Hamann (2016) suggest that the role of the academic *elite* may become even stronger through competitive funding (because the academic elites populate the selection panels) and through autonomy (because the academic elites occupy the managerial positions). This may in fact hinder the dynamics and innovation in the system and may result in lower efficiency and performance.
- A similar argument holds for the effects of stratified national science systems, of which Abramo et al. (2012) argues that these are a positive consequence of competition, and should result in higher performance. But one also may turn this argument upside down: in stratified systems, a few institutions acquire most of the reputation and through that most resources. This in turn weakens competition, as it does not create a performance based system but a reputation based system. It may produce and reinforce a few top universities at the expense of efficiency (UK) in an even stronger way than low competitive-high reputation systems do (Switzerland). In other words, we may face an *elitist circle*: competitive funding results in resource differences, which result in performance differences. The latter in turn reinforces the reputation differences. One ends with a reputation based funding system, and not in a performance based system. The *market* metaphor in relation to competitive funding may be the wrong one, as the system evolves to stable *power* positions. If this is correct, high competition is an unstable configuration, as it leads to high reputation based (locked in) systems, which not necessarily is the same as a high-performance system.

Overall, our findings do suggest that many of the generally accepted ideas about the relation between evaluation, competitive funding, university autonomy, academic freedom, and performance do not seem correct. That is not only an important research finding, it is also relevant for shaping science policy which should not listen to ideas that lack empirical support and may simply be irrelevant.

Funding

This work was supported by the Swedish "Riksbankens Jubileumsfond" (Tercentenary Foundation) through grant P12-1302:1.

Author contributions

Ulf Sandström: Conceived and designed the analysis; collected the data; contributed data or analysis tools; performed the analysis; wrote thepaper.

Peter van den Besselaar: Conceived and designed the analysis; collected the data; contributed data or analysis tools; performed the analysis; wrote thepaper.

Acknowledgements

The authors contributed equally to the paper. An earlier version was presented at STI 2014 conference in Leiden: Sandström, Heyman, and van den Besselaar (2014). The authors are grateful to 1) the editor and two anonymous reviewers

[5] However, also counter-forces do exist. For example, councils and university managers may try to do top-down agenda setting, but researchers are rather good in relabeling their research and remain doing what they (bottom-up) always do (Chapman & Farina, 1983; Dalpè and Anderson, 1995).

for their constructive comments; 2) Dr. Ulf Heyman for support with the collection and the harmonizing of the data (see Appendix A) and for his contribution in clarifying how to compare funding between countries.

Appendix A. Data

Coverage:

1. We started with all countries that around 2000 could be counted as developed science countries. So, the upcoming Asian and South American countries were not included, and the same counts for the East European countries.
2. We excluded the US from the analysis, because the US is too large to compare with the other countries. It would be worthwhile to in a next study include some or all US states separately, if information on that lower level of aggregation can be collected.
3. For Italy there is lack of information on project funding. Therefore, Italy is not included in Figs. 3–7. (Reliable) figures for Japan are only available for efficiency.

Efficiency:

1. We calculated over the period 2002–2011 the compound annual growth rate (CAGR) of the FP10% (field normalized top10% cited papers). We did the same operation for funding over the period 2000–2009. Figures are given in Table A1. Then we apply a linear regression over the set of countries (CAGR-output = a + b * CAGR-funding). The residual (measured growth rate minus expected growth rate) is the measure for the efficiency. CAGR was calculated using the LOGEST function in Excel: (LOGEST(ARRAY1,ARRAY2)-1)*100 where ARRAY1 are the values over a period and ARRAY2 the dates.

Table A1
Input and output data (CAGR) and residuals per country.

	Compund Annual Growth Rate (CAGR)		
Country	2000–09 Funding	2002–11 Pub(top10%)	Residuals
Australia	7.37	7.37	−0.52
Austria	4.20	3.66	−0.82
Belgium	2.87	2.87	2.59
Canada	4.42	4.42	−0.33
Denmark	4.45	4.61	0.38
Finland	3.70	3.70	−2.12
France	2.28	2.95	0.30
Germany	2.47	2.47	0.69
Ireland	11.42	11.42	−0.82
Israel	2.00	2.00	−2.10
Italy	3.66	3.94	1.64
Japan	2.02	2.60	−2.46
Netherlands	2.89	2.89	2.67
New Zealand	4.72	4.72	0.91
Norway	6.81	5.84	0.40
Spain	6.59	6.59	1.92
Sweden	2.66	2.64	−0.80
Switzerland	4.74	4.74	0.59
United Kingdom	4.93	4.93	−2.10

Input data:

1. The input data come from the OECD Science and Technology Indicators. We used the HERD for measuring input as the HE system produces most of the output. We did not use GOVERD because this also includes funding expenditures to R & D and Innovation (e.g. so-called defense research) which are not related to academic research – in an even less comparable way.
2. We use as indicator in this study the *change in funding*: the percentage funding increase/decrease over the period 2000–2009.
3. Portugal was removed, because a change in statistics made the estimate unreliable. Italy was removed from most of the calculations (e.g. institutional funding) as adequate data were not available (see Coverage, point 2 above).
4. In several other cases (Austria, France, Italy, Japan, Sweden), the input data were adapted to correct for changes in the statistics. For those countries we corrected for the break in the time series, by re-estimating historical data using the latest available definitions. The following corrections were applied: Austria: −0.54 (teacher's colleges); France: 0.68 (changed

statistics on staff time use); Italy: 2.60; Japan: 1.87; Norway: −0.97; Sweden: 0.15 (all due to changed statistics). Source: OECD MSTI Statistics. Related files. Documentation Appendix.
5. In countries where the structure of the system did change in the period under consideration, we adapted the HERD figures. This was the case for Denmark (correction −3.46), where the institutes sector was merged into the higher education sector during the period (which also affects the external funding figures, see Project funding point 5 below.).

Output:

1. We count output using the number of papers a country has in a year in the category top 10% best cited papers. We use fractional counting at country level.
2. The number of journals (in WoS) almost doubled in the period under consideration. This may influence the estimation of output growth. However, the growth mainly consisted of journals with a more regional/local focus. Therefore we used the CTWS core journal list.
3. What is not taken into account is the effect different national research portfolios have on the share of top 10% cited papers. However, elsewhere we showed that the developed countries do not have the same, but rather similar portfolios (Horlings & Van den Besselaar, 2013)
4. We use the percentage change in FP10% (fractionalized top 10%) over the period 2002–2011 to calculate efficiency change.
5. We use Web of Science subject categories for normalization of top 10% citations.

Project funding:

1. The share of project funding in total funding is based on a study of Van Steen, that used a systematic approach to make data between countries comparable (Van Steen, 2012; Versleijen et al., 2007). The figures reflect the 2008 situation but with necessary adaptions due to jumps in the statistics. E.g. the figure for Denmark has been adapted to the situation as it was before the institutes were merged with universities.
2. For some countries (France, Spain, UK) the Van Steen (2012) study lacks data. In these cases we used the OECD figures on GUF (General University Funds). The ratio between GUF and institutional funding for the countries covered by Van Steen (2012) is 1.27. The best guess we can make is multiplying the GUF of the three countries with that value in order to create better comparable data.
3. For Sweden (also not in the sample of Van Steen) we estimated the share of project funding from national statistics. Research funding to the academic hospitals were included to the general university funding.
4. For New Zealand, Van Steen reports 28.1% of total funding as institutional funding. This seems incorrect. We increased it to 59% based on the National Budget and S&T statistics.
5. Data on change in the share of project funding are not available for the period under consideration. The PREF study (Reale, 2017) suggest an increase, but mainly covers the period after 2008. The share of competitive institutional funds was rather stable. A few countries were reported in the PREF study with large changes. Inspecting the data, we found that this either may be a data problem (first two years in UK data), or the change started after the period under consideration in this study (Sweden). Or, as in Norway, it was a real change (Reale, 2017, p. 48 ff). We also found two countries without an NRES, but with reported high levels of competitive institutional funding (France, Israel). However, data for France cover the period after 2010, and for Israel, it also does not cover the period we study.

National research evaluation system (NRES):

1. We distinguish three classes of countries: those with a funding related national research evaluation system, those with a national evaluation system without funding implications, and those without a national research evaluation system.
2. The classification is based on a series of studies (see notes to Table 1).
3. Where the system was implemented after 2006, it is considered as to recent to have an effect on the efficiency in the 2000–2009 period.

Autonomy:

1. Data from the Esterman et al. (2011) study.

Academic freedom:

1. Data from the Teichler et al. (2013) study.

Stratification of the university system (Leiden ranking 2014):

1. Highest performing university in a national system
2. Performance differences within universities in a country

For each university $DF_{within} = (\text{StDev PP10\%}_{field}/\text{Mean PP10\%}_{field})$

For each country use the average of the DF_{within}

3. Performance differences between universities in a country

$$DF_{between} = (StDevPP10\%_{university}/MeanPP10\%_{university})$$

References

Abramo, G., & D'Angelo, C. A. (2017). The relationship among research productivity, research collaboration, and their determinants. *Journal of Informetrics*, *11*, 1016–1030.

Abramo, G., Cicero, T., & D'Angelo, C. A. (2012). The dispersion of research performance within and between universities as a potential indicator of the competitive intensity in higher education systems. *Journal of Informetrics*, *6*(155), 168.

Aghion, P., Dewatripont, M., Hoxby, C., Mas-Colell, A., & Sapir, A. (2007). *Higher aspirations; an agenda for reforming European universities*. Brussels: Breughel.

Aghion, P., Dewatripont, M., Hoxby, C., Mas-Colell, A., & Sapir, A. (2010). The governance and performance of universities: Evidence from Europe and the US. *Economic Policy*, 7–59.

Aksnes, D. W., Sivertsen, G., van Leeuwen, T. N., & Wendt, K. K. (2016). Measuring the productivity of national R&D systems: Challenges in cross-national comparisons of R&D input and publication output indicators. *Science and Public Policy*, *44*(2), 246–258.

Auranen, O., & Nieminen, M. (2010). University research funding and publication performance an international comparison. *Research Policy*, *39*(822), 834.

Bonaccorsi, A., & Daraio, C. (2004). Econometric approaches to the productivity of R & D systems. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research: The use of publication and patent statistics in studies of S&T systems.* (pp. 51–74). Dordrecht: Kluwer Acad. Publ.

Bonaccorsi, A. (2007). Explaining poor performance of European science: Institutions versus policies. *Science and Public Policy*, *34*, 303–316.

Bornmann, L., Stefaner, M., de Moya Anegon, F., & Mutz, R. (2014). What is the effect of country-specific characteristics on the research performance of scientific institutions? *Journal of Informetrics*, *8*(3), arXiv: 1401.2866v2

Chapman, I., & Farina, C. (1983). Peer review and national need. *Research Policy*, *12*, 317–327.

Cimini, G., Zaccaria, A., & Gabrielli, A. (2016). Investigating the interplay between fundamentals of national research systems: Performance, investments and international collaborations. *Journal of Informetrics*, *10*(1), 200–211.

Cole, S., & Phelan, T. J. (1999). The scientific productivity of nations. *Minerva*, *37*(1), 23.

Crespi, G. A., & Geuna, A. (2008). An empirical study of scientific production: A cross country analysis, 1981–2002. *Research Policy*, *37*, 565–579.

Dalpè, R., & Anderson, F. (1995). National priorities in academic research – Strategic research and contracts in renewable energies. *Research Policy*, *24*, 563–581.

Dasgupta, P., & David, P. (1994). Toward a new economics of science. *Research Policy*, *23*(5), 487–521.

Dawson, J., van Steen, J., & van der Meulen, B. (2009). *Science systems compared: A first description of governance innovations in six science systems*. Den Haag: Rathenau Instituut.

de Boer, H., Jongbloed, B., Benneworth, P., Cremonini, L., Kolster, R., Kottman, A., et al. (2015). *Performance-based funding and performance agreements in fourteen higher education systems.* [CHEPS report.]. www.utwente.nl/cheps

Esterman, T., Nokkala, T., & Steinel, M. (2011). *University autonomy in Europe II. The scorecard*. Brussels: European University Association.

Franzoni, C., Scellato, G., & Stephan, P. (2011). Changing incentives to publish (incl. Supporting material). *Science*, *333*(August), 702.

Garcia, C. E., & Sanz-Menéndez, L. (2005). Competition for funding as an indicator of research competitiveness. *Scientometrics*, *64*(3), 271–300.

Geuna, A., & Martin, B. R. (2003). University research evaluation and funding: An international comparison. *Minerva*, *41*, 277–304.

Gläser, J., & Laudel, G. (2016). Governing science how science policy shapes research content. *European Journal of Sociology*, *57*(1), 117–168.

Granberg, A., & Jacobsson, S. (2006). Myths or reality a scrutiny of dominant beliefs in the Swedish science policy debate. *Science and Public Policy*, *33*(321), 340.

Hagstrom, W. O. (1964). *The scientific community*. New York/London: Basic Books Inc.

Hall, R. (1972). In Nagi, & Corwin (Eds.), *Agencies of research support: Some sociological perspectives, the social contexts of research* (pp. 193–227). NY: Wiley-Interscience.

Hamann, J. (2016). The visible hand of performance assessment. *Higher Education*, *72*(6), 761–777.

Hazelkorn, E. (2011). *Rankings and the reshaping of higher education. The battle for world-class excellence*. New York: Palgrave Macmillan.

Heinze, T. (2008). How to sponsor ground-breaking research: A comparison of funding schemes. *Science and Public Policy*, *35*, 802–818.

Hicks, D. (2012). Performance-based university research funding systems. *Research Policy*, *41*(2), 251–262.

Horlings, E., & Van den Besselaar, P. (2013). *Convergence in science: Growth and structure of worldwide scientific output, 1993–2008*. The Hague: Rathenau Institute.

Hornbostel, S., Böhmer, S., Klingsporn, B., Neufield, J., & von Ins, M. (2009). Funding of young scientist and scientific excellence. *Scientometrics*, *79*(1), 171–190.

Jacobsson, S., & Rickne, A. (2004). How large is the Swedish 'academic' sector really? A critical analysis of the use of science and technology indicators. *Research Policy*, *33*(1355), 1372.

Jongbloed, B., & Vossensteyn, H. (2001). Keeping up performances: An international survey of performance-based funding in higher education. *Journal of Higher Education Policy and Management*, *23*(2), 127–145.

Jonkers, K., & Zacharewicz, T. (2015). *Performance based funding: a comparative assessment of their use and nature in EU member states*. JRC report. Brussels.

Latour, B., & Woolgar, S. (1979). *Laboratory life: The construction of scientific facts*. Princeton, New Jersey: Princeton Univ. Pr.

Laudel, G. (2006). The art of getting funded: How scientists adapt to their funding conditions. *Science and Public Policy*, *33*(7), 489–504.

Luukkonen, T. (2012). Conservatism and risk-taking in peer review: Emerging ERC practices. *Research Evaluation*, *21*, 48–60.

Luwel, M. (2004). Input data in the performance analysis of R & D systems. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research: The use of publication and patent statistics in studies of S&T systems* (pp. 315–338). Dordrecht: Kluwer Acad. Publ.

Merton, R. K. (1968). Matthew effect in science. *Science*, *159*, 56–63.

Moed, H. F., van Leeuwen, T. N., & Visser, M. S. (1999). Trends in publication output and impact of universities in the Netherlands. *Research Evaluation*, *8*(1), 60–67.

Musselin, C. (2013). How peer review empowers the academic profession and university managers: Changes in relationships between the state, universities and the professoriate. *Research Policy*, *42*, 1165–1173.

Nedeva, M., Boden, R., & Nugroho, Y. (2012). Rank and file: Managing individual performance in university research. *Higher Education Policy*, *25*(3), 335–360.

OECD. (2012). *OECD science, technology and industry outlook 2012*. OECD.

Orr, D., Jaeger, M., & Schwarzenberger, A. (2007). Performance-based funding as an instrument of competition in German higher education. *Journal of Higher Education Policy and Management*, *29*(1), 3–23.

Pan, R. K., Kaski, K., & Fortunato, S. (2012). World citation and collaboration networks: Uncovering the role of geography in science. *Scientific Reports*, *2*, 902. http://dx.doi.org/10.1038/srep00902

Paradeise, C., & Thoenig, J. C. (2013). Academic institutions in search of quality: Local orders and global standards. *Organization Studies*, *34*(2), 195–224.

Price, D. J. D. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, *27*(5), 292–306.
Reale, E. (2017). *Analysis of public research funding; JRC technical report*. Brussels: ERC.
Salmi, J. (2009). *The challenge of establishing world-class universities*. Washington, D.C: World Bank.
Sandström, U., & Hällsten, M. (2008). Persistent nepotism in peer review. *Scientometrics*, *74*(2), 175–189.
Sandström, U., & Heyman, U. (2015). In B. Fjaestad (Ed.), *What does funding mean for research quality?* Stockholm: Riksbankens jubileumsfond.
Sandström, U., & Van den Besselaar, P. (2016). Quantity and/or quality? The importance of publishing many papers. *PLoS One*, *11*(11), e0166149. http://dx.doi.org/10.1371/journal.pone.0166149
Sandström, U., & Wold, A. (2015). Centres of Excellence: Reward for gender or top-level research? In Björkman, & Fjaestad (Eds.), *Thinking ahead: research, funding and the future* (pp. 69–89). Stockholm: Makadam Publ.
Sandström, U., Heyman, U., & van den Besselaar, P. (2014). The complex relationship between competitive funding and performance. In Noyons (Ed.), *Context counts: Pathways to master big and little data - STI 2014* (pp. 523–533). Leiden: CWTS.
Schubert, T. (2009). Empirical observations on new public management to increase efficiency in public research—Boon or bane? *Research Policy*, *38*(8), 1225–1234.
Senker, J. (1999). *Final report: European comparison of public research systems*. TSER Project no SOE1–CT96–1036, SPRU report, September.
Stephan, P. E. (2012). *How economics shape science*. Harvard U.P.
Teichler, U., Arimoto, A., & Cummings, W. K. (2013). *The changing academic profession*. Dordrecht etc: Springer.
Thomas, D., & Nedeva, M. (2012). Characterizing researchers to study research funding agency impacts: The case of the European Research Council's Starting Grants. *Research Evaluation*, *21*(4), 257–269.
Tijssen, R. J. W., Visser, M. S., & van Leeuwen, T. N. (2002). Benchmarking international scientific excellence: Are highly cited research papers an appropriate frame of reference? *Scientometrics*, *54*(3), 381–397.
Van Arensbergen, P., van der Weijden, I., & van den Besselaar, P. (2014). Different views on scholarly talent – What are the talents we are looking for in science? *Research Evaluation*, *23*(4), 273–284.
Van Steen, J., & Eijffinger, M. (1998). Evaluation practices of scientific research in the Netherlands. *Research Evaluation*, *7*(2), 113–122.
Van Steen, J. (2012). *Modes of public funding of research and development: Towards internationally comparable indicators*. OECD, Science Technology and Industry. Working Papers, 2012/04.
Van den Besselaar, P., & Leydesdorff, L. (2009). Past performance, peer review, and project selection: A case study in the social and behavioral sciences. *Research Evaluation*, *18*(4), 273–288.
Van den Besselaar, P., & Sandström, U. (2015). Early career grants, performance, and careers: A study on predictive validity of grant decisions. *Journal of Informetrics*, *9*, 826–838.
Van der Weijden, I., Verbree, M., & Van den Besselaar, P. (2012). From Bench to Bedside: The societal orientation of research leaders. The case of biomedical and health research in the Netherlands. *Science & Public Policy*, *39*, 285–303.
Versleijen, A., Van der Meulen, B., Van Steen, J., Kloprogge, P., Braam, R., Mamphuis, R., et al. (2007). *Dertig jaar onderzoeksfinanciering – Trends, beleid en implicaties (Thirty years research funding in the Netherlands)*. Den Haag: Rathenau Instituut.
Waltman, L., & Schreiber, M. (2013). On the calculation of percentile-based bibliometric indicators. *JASIST Journal of the American Society for Information Science and Technology*, *64*(2), 372–379.
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. *JASIST Journal of the American Society for Information Science and Technology*, *63*(12), 2419–2432.
Wennerås, C., & Wold, A. (1997). Nepotism and sexism in peer-review. *Nature*, *387*(22 May), 341–343. http://dx.doi.org/10.1038/387341a0
Westerheijden, D. F. (1997). A solid base for decisions: Use of VSNU research evaluations in Dutch universities. *Higher Education*, *33*(4), 397–413. http://dx.doi.org/10.1023/A:1002995623272
Whitley, R. (2007). Changing governance of the public sciences: The consequences of establishing diverse research evaluation systems. In R. Whitley, & J. Glaeser (Eds.), *The changing governance of the sciences: The advent of research evaluation systems*. Dordrecht, Netherlands: Springer, p. 3-27 25 p.