Benefit of the doubt approach to assessing the research performance of Australian universities

Andrzej Szuwarzyński
Faculty of Management and Economics, Gdańsk University of Technology, Gdańsk, Poland

Abstract
This paper proposes a benefit of the doubt (BoD) approach to assess the research performance of 37 public Australian universities based on data from 2015. The primary activities of higher education institutions are teaching and research, but the prestige of a university depends mainly on the results of its research activities. The BoD method is rooted in the data envelopment analysis methodology, which enables the flexible and data-based assignment of weights to aggregated variables. Full weighting flexibility, however, allows zero weights, which can lead to unrealistic results. For this reason, the proposed model has been supplemented with additional weight restrictions. The assessment considers key performance factors: number of publications and citations; number of completed doctoral degrees; amount of research grants; and percentage of science graduates. Unlike earlier research on Australian universities, this study uses the numbers of publications and citations from the Web of Science database. The results provide a ranking of universities and recommendations for decision makers regarding the direction of future improvement actions for worse-performing universities.
Keywords:
research performance, research activities, performance factors, weight restrictions, benefit of the doubt, data envelopment analysis

Introduction

The primary areas of university activity are teaching students and conducting scientific research. The prestige of a university and its position in the competitive higher education market depend, however, mainly on the results of the university's research activities. This fact is reflected in the global rankings of universities. The famous and prestigious Shanghai Academic Ranking of World Universities (ARWU) focuses on research activities. The Times Higher Education (THE) ranking assesses universities according to their scientific findings, elements of the internationalisation of teaching and research, and subjective assessments, including employers' opinions. The rankings are built based on composite indicators (CIs) that aggregate a set of variables using weights that reflect the importance of individual variables (Paruolo, Saisana, & Saltelli, 2013). However, the use of this aggregation method has frequently been criticised due to both the choice of variables used and the weighting scheme adopted (Marginson, 2007, 2014; Paruolo et al., 2013). Numerous studies have also presented national rankings based mostly on non-parametric methods (see, for example, Abbott & Doucouliagos, 2003; Carrington, Coelli, & Rao, 2005; Johnes, 2006; Katharaki & Katharakis, 2010; Lee, 2011; Worthington & Lee, 2008).

However, despite these objections, CIs are widely used. Rankings based on CIs are employed to benchmark the performance of universities and higher education systems (De Witte & Rogge, 2010), and CIs help decision makers set policy priorities. Compiling multiple individual indicators that describe complex problems into a single index facilitates interpretation by decision makers, the media and the general public (Cherchye, Moesen, Rogge, & Van Puyenbroeck, 2011; Shen et al., 2011). Individual universities are also increasingly interested in evaluating the work of their staff, in terms of both teaching efficiency and research achievement, because the obtained results can be used in their staffing policies (De Witte & Rogge, 2010). In addition to decision makers in higher education systems, rankings are of interest to students selecting a university and to employers who consider the rankings when recruiting employees, and other stakeholders use these rankings to make decisions related to, e.g., funding or sponsorship. Despite the criticism of rankings based on CIs, they are likely the only widely available benchmark for higher education institutions (HEIs) (Gnaldi & Ranalli, 2016; Marconi & Ritzen, 2015).

The key issue is the selection of variables for CIs. The basic outcome of a university is increased knowledge, which can be divided into two types: person-specific (e.g., undergraduate and postgraduate students) and general (e.g., books, publications, patents) (Beasley, 1990). HEIs produce knowledge through research and transmit it through the formation of graduates and the publication of research results (Pastor, Serrano, & Zaera, 2015), and research activity is recognised as the core activity of universities (Agasisti, Dal Bianco, Landoni, Sala, & Salerno, 2011). Measurement of research results in HEIs is problematic because of the multi-product nature of educational and research activities (Pastor et al., 2015). These activities are far too complex to be grasped by a single output criterion (De Witte & Rogge, 2010).

Despite the widespread use of CIs for the creation of rankings, they are controversial—mainly because of the weighting schemes for individual sub-indicators. For example, for the ARWU and THE rankings, weights are determined subjectively based on expert judgement, and it is common practice to assume equal weights for all aggregated sub-indicators (see, for example, Despotis, 2005; Manca, Governatori, & Mascherini, 2010).

This paper presents a model for measuring and comparing the research activities of Australian public universities. The purpose of this paper is to build a model to assess the performance of research activities and to formulate recommendations for the worse performing units regarding the directions and magnitude of the changes needed to improve efficiency, based on the example of Australian public universities. As a starting point, a common approach based on CIs is adopted. A critical evaluation of the standard model of CIs is conducted, and an extension of this methodology is introduced, contributing to improving the overall credibility of the results of performance assessment. Three variants are presented, of which two are based on the benefit of the doubt (BoD) approach. The final model is supplemented with two additional options to minimise the basic shortcomings of the BoD approach, such as the acceptability of zero weights during the aggregation of sub-indicators and the inability to create a ranking of fully efficient units. For this purpose, additional weight restrictions are introduced, and a super-efficiency algorithm is used. Unlike earlier research on Australian universities, the numbers of publications and citations from the Web of Science (WoS) database are used in this study. The proposed methodology is illustrated by an empirical application using data collected on 37 Australian public universities in 2015. The calculations are performed using MaxDEA Ultra software, version 6.18.

The main practical implication of this study is that the conclusions based on its results can help decision makers evaluate the research activities of universities and contribute to identifying the directions of future improvement actions.

The remainder of this paper is organised as follows. The second section provides a brief overview of the state of the art regarding the choice of inputs and outputs in university efficiency studies, particularly in Australia. The third section describes the methodology used to construct CIs using the BoD approach with the aforementioned extensions. The fourth section contains the justifications and description of the factors included in the model, and the final section discusses the results obtained from the three models. The paper ends with brief concluding remarks.

The choice of inputs and outputs in university efficiency studies

Studies on the efficiency of Australian universities have been conducted in recent years by many authors, and these studies have concerned both combined teaching and research activities (Abbott & Doucouliagos, 2003; Avkiran, 2001; Carrington et al., 2005; Carrington, O'Donnell, & Rao, 2016; Worthington & Lee, 2008) and research activity itself (Abbott & Doucouliagos, 2004; Lee, 2011; Valadkhani & Worthington, 2006). Similar studies have also been conducted in other countries, e.g., the United Kingdom (Johnes, 2006), Greece (Katharaki & Katharakis, 2010), Italy (Agasisti et al., 2011) and Spain (Murias, de Miguel, & Rodriguez, 2008). The aforementioned examples of studies refer to comparisons of universities or units operating within universities. Different models of the non-parametric data envelopment analysis (DEA) method were used to evaluate their efficiency. DEA is a data-oriented approach for evaluating the performance of a set of peer entities called decision making units (DMUs), which convert multiple inputs into multiple outputs. DEA models measure the efficiency of a DMU relative to similar DMUs in order to estimate a 'best practice frontier' (Cooper, Seiford, & Zhu, 2011). Such models could be represented by a linear programming technique in which each DMU tries to maximise the efficiency ratio endogenously by determining the best set of weights (Agasisti et al., 2011).

Depending on the purpose of the research, a set of inputs and outputs describing the performance of the university is determined. DEA assumes that inputs are generally the resources needed to produce outputs. The inputs used in the efficiency analysis usually include costs, academic and non-academic full-time equivalent (FTE) staff and the numbers of undergraduate, postgraduate and doctorate students. The basic teaching outputs generally include the number of graduates, while in the case of research, outputs such as the number of publications, the number of citations, the number of research projects or the income from research are typically considered (see, for example, Johnes, 2006; Katharaki & Katharakis, 2010; Worthington & Lee, 2008).

Studies on Australian universities' research activity have considered only 'research only' and 'teaching and research' academic staff (Lee, 2011; Valadkhani & Worthington, 2006). It is assumed that 'teaching only' academic staff are not involved in research. Graduates are divided according to two criteria. The first is a division into undergraduate, postgraduate and doctoral graduates. The second is a division into science students and non-science students. This division is due to the varied intensity of resource use in the teaching process. Science students include disciplines such as natural and physical sciences, information technology, engineering and related technologies, architecture and building, agriculture, environmental and related studies, and health (Carrington et al., 2016).

The most serious controversies are caused by factors such as research income and publications. There is no agreement regarding whether research income should be treated as an input or an output. According to Beasley (1990), research income is used as a measure of output. Other authors (Abbott & Doucouliagos, 2004; Agasisti et al., 2011; Agasisti & Haelermans, 2016; Katharaki & Katharakis, 2010) share the same view and have stated that research income flows into universities as a result of the research work of their academic staff. Consequently, research income can be treated as an output because it reflects the market value of university research output. However, there is another approach to interpreting research income—treat research income as research input because it is used to hire staff and to purchase the materials and equipment needed to perform research (Carrington et al., 2016). In this article, research income is treated as an output, as in (Agasisti et al., 2011; Agasisti & Haelermans, 2016). This approach is also supported by the fact that a large part of the research activity is performed by 'teaching and research' academic staff, combining teaching with research activities. It is not possible to determine to what extent research activities are financed from research or from teaching.

Research activity data from the WoS database provide relatively homogeneous, global information on university publications and citations (Docampo & Cram, 2017; Lukman, Krajnc, & Glavič, 2010). However, it should be noted that other practices have been applied in prior research performance studies on Australian universities. The previous studies used the weighted publications index (Carrington et al., 2005; Carrington et al., 2016), which is a measure the Commonwealth employs to help allocate research funds to universities. This measure includes books, book chapters, journal articles and conference papers. Books are assigned the highest weight of five, and the remaining research categories receive an equal weight of one. The Australian Research Council (2015) uses the weighted publications index to measure university research output. Clearly, not all university research outcomes are reflected in this measure, and it does not take into account quality and the influence that research has on contributing to increased knowledge or the prestige of a journal (Carrington et al., 2016). The weighted publications index is correlated with the number of publications in WoS at the level of 0.98 (based on 2015 data). The use of WoS data, apart from the number of publications, allows the impact of these publications to be considered through the number of citations as a proxy for the quality of published research outcomes.

An analysis of the available literature indicates that in surveys of university research activity, individual authors use similar sets of outputs. Pastor et al. (2015) stated that the measurement of the research outcomes of HEIs is problematic due to the multi-product nature of university activities. HEIs produce knowledge through research and transmit it via the production of graduates and the publication of research results. In these authors' opinion, not only the quantity but also the quality of research output is important. Finally, the authors proposed the number of scientific publications as research output and the number of citations per document as the quality parameter of this output. It can be stated that the research activities of universities constitute the basis for the development of a knowledge society and impact socioeconomic development. Marconi and Ritzen (2015) noted that indicators should be related to the mission of a university. The underlying assumption is that a research-oriented university assigns high priority to PhD studies. In their study, these authors use the ratio between PhD students and total students at a university as a proxy for how research-oriented an institution is. In addition to this variable, the authors use the number of articles published and the number of graduates in the natural sciences. The authors also suggest using the number of professors or computers as an input. Similarly to other authors, Barra and Zotti (2016) adopted total external research funding obtained by the university and the weighted sum of scientific publications as research outputs. These authors also adopted the total number of academic staff and non-academic staff and the total amount of financial resources spent on research activities as research inputs. Valadkhani and Worthington (2006) concluded that research income, the number of academic staff and the number of postgraduates are positively correlated with research output. The authors defined three measures of research output in their analysis: the number of PhD completions, the number of weighted publications and the total amount of research grants.

Research methodology

CIs enable aggregation of many sub-indicators into one measure, making it possible to compare many objects. CIs also integrate large amounts of information in a clear and understandable format that is easy for recipients to interpret (Shen et al., 2011). However, CIs are criticised as possibly misleading, particularly when used to rank complex economic phenomena. CIs present some methodological difficulties that must be confronted, and they can be easily manipulated to produce desired outcomes (Freudenberg, 2003). Weights are usually adopted on the basis of expert opinion, but identical weights for all factors are adopted in many applications (see, for example, Despotis, 2005; Manca et al., 2010). Solutions such as equal weighting or invoking the opinion of experts in the thematic field under scrutiny are therefore often used, although it is telling that diverging opinions are regularly put forward (Van Puyenbroeck, 2017). Cherchye, Moesen, Rogge, and Van Puyenbroeck (2007) have a similar opinion—these authors claim that despite increasing use of CIs, they remain controversial. The undesirable dependence of rankings on the preliminary normalisation stage and the disagreement among experts on the weighting scheme used to aggregate sub-indicators are often invoked to question the credibility of CIs. Despite these drawbacks, CIs are applied in both policy shaping and operational decision making, with the Human Development Index (Despotis, 2005) and the model of technological capabilities assessment (Filippetti & Peyrache, 2011) as examples of their use. The methodology of creating CIs was disseminated by the Organisation for Economic Co-Operation and Development (2008). Based on the linear rule of aggregation of normalised sub-indicators, the CI for object *j* is defined as:

$$CI_j = \sum_{i=1}^{m} w_i I_{ij} \quad \text{for } j = 1, \dots, n$$

$$\text{subject to} \quad \sum_i w_i = 1 \text{ and } 0 \le w_i \le 1, \quad \text{for } i = 1, \dots, m \qquad (1)$$

where CI_j is the value of the CI for object *j*; w_i is the weight of sub-indicator *i*; I_{ij} is the value of the normalised sub-indicator *i* for object *j*; *n* is the number of objects incorporated into the analysis; and *m* is the number of sub-indicators.

The basic shortcoming of the standard aggregation method is the arbitrary weighing scheme that assumes, for each of the sub-indicators, equal weights for all objects (Lovell, Pastor, & Turner, 1995; Moesen & Cherchye, 1998). However, the BoD method eliminates the dependency of the results on preliminary normalisation, and the method's characteristic of offering flexibility in the form of endogenous weighting may reduce some of the aforementioned criticisms of CIs. DEA may be instrumental in overcoming these limitations. One part of the appeal of DEA in the context of CIs stems from its invariance to measurement units, which entails the possibility of skipping the normalisation stage (Cherchye et al., 2007). The precursors of this method were Melyn and Moesen, who introduced a synthetic measure of macroeconomic performance, the Leuven Index of Macroeconomic Performance (LIMEP), in 1991. The technique used to calculate the LIMEP was inspired by DEA (Moesen & Cherchye, 1998). The DEA method in its classical form has a strong connection with the theory of production in economics. From the most general perspective, DEA minimises 'inputs' and maximises 'outputs'; in other words, lower levels of the former and higher levels of the latter represent better productivity or efficiency. The BoD approach allows for benchmarking, in which best performing objects form the 'best practice frontier', assuming that all of the factors included in the model are outputs. In such a situation, the outputs, as measures of performance, must fulfil the condition of 'the more, the better' (Cook, Tone & Zhu, 2014).

The use of DEA to construct CIs has been popularised by Cherchye et al. (2007). This approach, known as BoD CI construction, is equivalent to the input-oriented DEA model, assuming constant returns to scale, and was proposed by Charnes, Cooper, and Rhodes (1978). The basic difference between the original DEA model and the BoD approach consists of CIs usually looking for 'achievements' without considering the input side (Cherchye et al., 2007). As a result, all of the sub-indicators are considered to be outputs, and the only input is the dummy variable equal to one for all objects. In this sense, the dummy input for each object can be interpreted as a 'helmsman' (Koopmans, 1951), which accomplishes certain goals corresponding to different sub-indicators (Lovell et al., 1995; Murias et al., 2008). To obtain a ranking of fully efficient objects, the basic model was extended by an additional constraint, allowing the comparison of the evaluated object with a linear combination of all of the other objects in the sample by excluding the object from the evaluation; the so-called super-efficiency model (Andersen & Petersen, 1993; Malul, Hadad, & Ben-Yair, 2009). The problem of maximisation, i.e., input oriented under constant returns to scale, can be written for each object k in linear form as:

$$CI_k = \max_{w_{ik}} \sum_{i=1}^{m} w_{ik} y_{ik}$$

$$\text{subject to} \quad \sum_{i=1}^{m} w_{ik} y_{ij} \leq 1 \quad \text{for } j = 1, \ldots, n, \qquad j \neq k \tag{2}$$

$$w_{ik} \geq 0 \quad \text{for } i = 1, \ldots, m$$

where CI_k is the value of the CI for object k; w_{ik} is the weight of sub-indicator i for object k; y_{ik} is the value of sub-indicator i for object k; n is the number of objects incorporated into the analysis; and m is the number of sub-indicators.

One well-known feature of DEA is that it seeks endogenous weights that maximise the overall score for each object given a set of other observations (Cherchye et al., 2007). Although DEA allows the optimal weights of indicators to be determined, in some cases, it might be important to include expert opinions on weight restrictions in the model. As stated by Cherchye et al. (2011), the possibility of adding information related to the relevance of individual sub-indicators allows for increased credibility and the acceptance of CIs in practical applications. The full flexibility of determining weights provided by DEA can, in some situations, allow an object to show itself to be a brilliant performer in a difficult to justify fashion, e.g., if some of the calculated weights are zero, because global performance is based on a small subset of sub-indicators (Cherchye et al., 2007; Podinovski & Thanassoulis, 2007; Cooper, Ruiz, & Sirvent, 2009; Rogge 2011).

To avoid this shortcoming, weight restrictions should be added to equation (2) (Angulo-Meza & Lins, 2002; Cherchye et al., 2007; Mecit & Alp, 2013). In this paper, the virtual weight restrictions first proposed by Wong and Beasley (1990) are used. The below restrictions are entered into model (2) for each output (Allen, Athanassopoulos, Dyson, & Thanassoulis, 1997; Angulo-Meza & Lins, 2002; Zanella, Camanho, & Dias, 2015):

$$\alpha_i \leq \frac{w_i y_{ij}}{\sum_{i=1}^{m} w_i y_{ij}} \leq \beta_i \qquad (3)$$

where α_i and β_i are the lower and upper bounds, respectively, for output i.

Specification of $[\alpha_i, \beta_i]$ is a value judgement. Such judgements indicate that according to a decision maker's opinion, the model represents the modelled phenomenon better due to imposing such restrictions (Wong & Beasley, 1990). Introducing weight restrictions improves discrimination (Angulo-Meza & Lins, 2002). The application of weight restrictions requires running the classic model without restrictions to determine the weight dimension for each of the outputs and implementing restrictions on this basis. If the results of the model with restrictions prove infeasible, the restrictions should be relaxed until the infeasibility disappears (Angulo-Meza & Lins, 2002). According to Sarrico and Dyson (2004), the imposition of weight restrictions on virtual outputs requires the application of a rather output-oriented model. The values of CIs can also be obtained from an output-oriented model because, for models with constant returns to scale, the efficiency scores obtained with different model orientations are the same (Van Puyenbroeck, 2017; Zanella et al., 2015).

Description of the data

This article uses data from the Australian Department of Education and Training website (DET, 2017), from which all of the data on the functioning of the Australian higher education system regarding teaching and research processes and their funding are available. The numbers of publications and citations are derived from the WoS database (WoS, 2017). Data from 37 public Australian universities for 2015 were used. The list of universities is shown in Table 1.

Table 1. List of universities

University name†	Abbreviation†	University name†	Abbreviation†
Australian Catholic University	ACU	Swinburne University of Technology	SWINBURNE
Australian National University	ANU	University of Adelaide	ADELAIDE
Central Queensland University	CQU	University of Canberra	CANBERRA
Charles Darwin University	CDU	University of Melbourne	MELBOURNE
Charles Sturt University	CSU	University of New England	UNE
Curtin University of Technology	CURTIN	University of New South Wales	UNSW
Deakin University	DEAKIN	University of Newcastle	NEWCASTLE
Edith Cowan University	ECU	University of Queensland	UQ
Federation University‡	FEDUNI	University of South Australia	UNISA
Flinders University of South Australia	FLINDERS	University of Southern Queensland	USQ
Griffith University	GRIFFITH	University of Sydney	SYDNEY
James Cook University	JCU	University of Tasmania	UTAS
La Trobe University	LATROBE	University of Technology Sydney	UTS
Macquarie University	MACQUAIRIE	University of the Sunshine Coast	USC
Monash University	MONASH	University of Western Australia	UWA
Murdoch University	MURDOCH	University of Wollongong	UOW
Queensland University of Technology	QUT	Victoria University	VU
RMIT University	RMIT	Western Sydney University	UWS
Southern Cross University	SCU		

† *- Names and abbreviations from http://www.australianuniversities.com.au/list/*
‡ *- Previously University of Ballarat*

The analysis of the literature performed in section 2 allowed the selection of a set of raw data characterising 10 factors relevant to the research activity of universities. The factors included in the analysis were as follows: ST_R—the number of FTE research-only academic staff; ST_T_R—the number of FTE teaching and research academic staff; SE_L—the number of academic staff in the ‘senior lecturer’ position; A_SE_L—the number of academic staff in positions above ‘senior lecturer’; DOC—the number of completed ‘doctorates by research’; T_GRAD—the total number of graduates; S_GRAD—the number of graduates in science disciplines; PUBL—the number of publications; CIT—the number of citations; and GRANT—the total amount of research grants.

The above factors were then converted into five ratios in accordance with Sagarra, Mar-Molinero, and Agasisti (2017) and Marconi and Ritzen (2015), who suggest that such a transformation is necessary to make the results independent of university size so that the level of raw outputs is relative to the size of the university.

Five sub-indicators were included for further analysis. O1=DOC/ (SE_L+A_SE_L)—the number of completed ‘doctorates by research’ per academic authorised to supervise doctoral dissertations (usually senior lecturers and above). This indicator illustrates the level of use of academic staff with the highest qualifications for supervising doctoral dissertations based on research. O2=S_GRAD/ T_GRAD—share of graduates in science disciplines among the total number of graduates. This is a proxy for universities’ involvement in the more research-related fields of study. O3=PUBL/ (ST_R+ST_T_R)—the number of publications per ‘research only’ and ‘teaching and research’ academic staff member required to conduct research. This is a proxy for the productivity of the research work conducted by academic staff. O4=CIT/ (ST_R+ST_T_R)—the number of citations per ‘research only’ and ‘teaching and research’ academic staff member. This is a proxy for publication quality. O5=GRANT/ (ST_R+ST_T_R)—the amount of obtained grants per ‘research only’ and ‘teaching and research’ academic staff member (in thousands of Australian dollars). This is a proxy for the market value of the conducted research.

Results and discussion

The calculations were performed using three methods based on the same set of data. The results of the first method, described by equation (1) are shown in Table 2. This result is a composite index calculated from the 5 sub-indicators that were previously normalised with the min-max method. The adopted aggregation scheme assumes equal weights for all the sub-indicators. The ‘Score’ column lists the values of the calculated indicator, and the ‘Rank’ column lists the ranks of the universities. The last five columns (O1-O5) contain all the normalised values for the sub-indicators.

Table 2. Australian university ranking calculated using CIs and the normalised values of the sub-indicators (in descending order)

University	Score	Rank	Normalised values of sub-indicators				
			O1	O2	O3	O4	O5
SYDNEY	0.821	1	0.606	0.771	1.000	1.000	0.728
UWA	0.782	2	0.581	1.000	0.894	0.792	0.643

MELBOURNE	0.780	3	0.675	0.641	0.910	0.901	0.775
MONASH	0.749	4	1.000	0.523	0.811	0.717	0.697
UQ	0.738	5	0.925	0.865	0.663	0.611	0.624
ADELAIDE	0.682	6	0.711	0.769	0.683	0.705	0.542
UNSW	0.610	7	0.548	0.718	0.661	0.564	0.560
CURTIN	0.609	8	0.656	0.699	0.691	0.591	0.409
ANU	0.582	9	0.691	0.131	0.662	0.581	0.847
UTAS	0.573	10	0.512	0.883	0.554	0.329	0.587
NEWCASTLE	0.524	11	0.659	0.635	0.559	0.367	0.399
CDU	0.510	12	0.323	0.489	0.446	0.294	1.000
JCU	0.480	13	0.400	0.729	0.468	0.502	0.303
UNISA	0.475	14	0.533	0.686	0.481	0.379	0.295
LATROBE	0.465	15	0.709	0.670	0.409	0.207	0.330
FLINDERS	0.438	16	0.361	0.724	0.476	0.378	0.252
UOW	0.427	17	0.588	0.479	0.497	0.360	0.211
QUT	0.413	18	0.587	0.515	0.394	0.272	0.296
MACQUAIRIE	0.411	19	0.852	0.000	0.535	0.405	0.262
GRIFFITH	0.408	20	0.368	0.418	0.577	0.422	0.253
DEAKIN	0.396	21	0.470	0.539	0.519	0.338	0.116
UTS	0.387	22	0.422	0.515	0.504	0.292	0.202
SCU	0.317	23	0.480	0.240	0.471	0.182	0.211
SWINBURNE	0.294	24	0.525	0.409	0.216	0.218	0.104
ECU	0.292	25	0.536	0.439	0.248	0.141	0.098
MURDOCH	0.284	26	0.448	0.196	0.404	0.217	0.153
USC	0.275	27	0.156	0.572	0.314	0.167	0.167
UWS	0.270	28	0.229	0.430	0.404	0.226	0.059
RMIT	0.266	29	0.510	0.318	0.211	0.108	0.183
UNE	0.266	30	0.471	0.256	0.251	0.085	0.266
CANBERRA	0.248	31	0.360	0.213	0.384	0.177	0.104
ACU	0.226	32	0.000	0.520	0.390	0.122	0.096
CQU	0.209	33	0.034	0.568	0.251	0.109	0.082
CSU	0.196	34	0.398	0.250	0.233	0.049	0.050
FEDUNI	0.176	35	0.115	0.488	0.220	0.058	0.000
VU	0.173	36	0.636	0.114	0.000	0.000	0.116
USQ	0.159	37	0.243	0.369	0.078	0.030	0.074

The main practical use of CIs is the creation of rankings. However, it is also possible to perform qualitative assessments of the strengths and weaknesses of individual universities on the basis of the normalised value of sub-indicators. The analysis presented below concerns the six highest-ranked and six lowest-ranked universities.

The top seven universities belong to the Group of Eight (Go8). The last university from this group (ANU) is at rank nine, while rank eight is taken by CURTIN, which belongs to the Australian University Technology Network (ATN). Therefore, the results are consistent with the generally accepted classification of research universities in Australia. The best universities are characterised by balanced development in all the areas described by the five sub-indicators. For the top six universities, the minimum value of all the sub-indicators is 0.523, with the average of all sub-indicators equal to 0.759. The best university in this ranking, SYDNEY, is the leader in two of the sub-indicators, O3 and O4, measuring the numbers of publications and citations, respectively. UWA leads in the number of science graduates (sub-indicator O2), while MONASH has the best sub-indicator score for 'doctorates by research' per promoter (sub-indicator O1). CDU, which belongs to the Innovative Research Universities (IRU) group, is twelfth in the ranking. CDU ranks highest in obtaining grants (sub-indicator O5); however, in the case of the remaining four areas, it has sub-indicators ranging from 0.294 to 0.489.

For the six universities at the bottom of the ranking, the minimum value of all of the sub-indicators is 0.000, with an average of all the sub-indicators equal to 0.190. Compared to the leaders, in these universities, specific areas are clearly mutually different. ACU has an O1 sub-indicator score equal to zero; however, its O2 sub-indicator score is 0.520, which is close to the score of one of the leaders, MONASH. FEDUNI has a zero value for O5, which is a sign of low interest in obtaining grants, but its O2 sub-indicator score is 0.488. Consequently, these two universities focus more on teaching activities in the science disciplines. In contrast, VU has zero values for O3 and O4, indicating that it pays little attention to publishing in highly reputed journals; however, the value of its O1 sub-indicator is relatively high (0.636), indicating that it strongly emphasises awarding 'doctorates by research', which could be treated as a contradiction. MACQUAIRIE has a zero value for the O2 sub-indicator, placing it in nineteenth position and indicating that it shows little interest in teaching the science disciplines; however, MACQUAIRIE has a relatively high value of 0.852 for sub-indicator O1, indicating that it pays significant attention to awarding 'doctorates by research'. Additionally, MACQUAIRIE's values for the O3 sub-indicator (0.535) and the O4 sub-indicator (0.405) indicate that it pays attention to publishing in renowned journals.

The universities included in this study have diverse activity profiles. A university's position at the bottom of the ranking does not prejudge it with a lower value. The purpose of the study is to evaluate research activities—it is important to remember that some of the universities are typically teaching in nature.

Table 3. Australian university rankings calculated using the BoD approach and the non-normalised values of the sub-indicators (in descending order according to BoD-R)

University	BoD		BoD-R		Non-normalised values of sub-indicators				
	Score	Rank	Score	Rank	O1	O2	O3	O4	O5
MONASH	1.108	4	1.102	1	0.57	0.38	3.30	11.01	13.20
UQ	1.108	3	1.095	2	0.53	0.52	2.82	9.59	12.00
SYDNEY	1.104	5	1.091	3	0.39	0.48	3.92	14.81	13.71
UWA	1.130	2	1.068	4	0.38	0.57	3.58	12.02	12.31
CDU	1.197	1	1.053	5	0.26	0.37	2.11	5.34	18.15
MELBOURNE	1.024	6	1.023	6	0.42	0.43	3.63	13.48	14.47
ANU	1.016	7	1.007	7	0.43	0.22	2.82	9.19	15.64
ADELAIDE	0.936	8	0.933	8	0.44	0.48	2.89	10.85	10.66
UTAS	0.921	9	0.907	9	0.35	0.53	2.46	5.81	11.40
CURTIN	0.883	10	0.880	10	0.41	0.45	2.91	9.33	8.49
UNSW	0.861	12	0.861	11	0.36	0.46	2.81	8.97	10.95
MACQUAIRIE	0.882	11	0.833	12	0.50	0.17	2.40	6.83	6.08
LATROBE	0.841	13	0.823	13	0.43	0.44	1.99	4.18	7.19
NEWCASTLE	0.821	14	0.817	14	0.41	0.43	2.48	6.32	8.32
UNISA	0.811	15	0.786	15	0.36	0.45	2.22	6.49	6.62
JCU	0.810	16	0.774	16	0.30	0.46	2.18	8.14	6.76
FLINDERS	0.807	17	0.747	17	0.28	0.46	2.21	6.47	5.93
UOW	0.740	18	0.732	18	0.38	0.36	2.28	6.23	5.25
QUT	0.725	19	0.717	19	0.38	0.38	1.94	5.04	6.64
DEAKIN	0.722	20	0.710	20	0.33	0.39	2.35	5.93	3.70
UTS	0.695	24	0.687	21	0.31	0.38	2.30	5.32	5.10
GRIFFITH	0.684	25	0.678	22	0.28	0.34	2.54	7.06	5.94
ECU	0.670	26	0.651	23	0.36	0.35	1.46	3.29	3.41
SWINBURNE	0.658	28	0.642	24	0.35	0.34	1.35	4.33	3.49
SCU	0.640	30	0.630	25	0.33	0.27	2.19	3.84	5.25
RMIT	0.634	31	0.618	26	0.34	0.30	1.34	2.85	4.80
VU	0.709	21	0.614	27	0.40	0.22	0.65	1.40	3.69
ACU	0.664	27	0.589	28	0.11	0.38	1.93	3.03	3.37
MURDOCH	0.594	34	0.587	29	0.32	0.25	1.97	4.31	4.30
UNE	0.598	33	0.586	30	0.33	0.27	1.47	2.54	6.14
USC	0.701	22	0.586	31	0.18	0.40	1.68	3.64	4.53
UWS	0.601	32	0.575	32	0.22	0.34	1.97	4.43	2.76
CANBERRA	0.562	36	0.552	33	0.28	0.26	1.91	3.77	3.50
CQU	0.698	23	0.535	34	0.13	0.40	1.47	2.87	3.14
CSU	0.548	37	0.530	35	0.29	0.27	1.41	2.05	2.63
FEDUNI	0.642	29	0.506	36	0.17	0.37	1.37	2.18	1.80
USQ	0.565	35	0.505	37	0.22	0.32	0.90	1.80	3.00

Further results are presented in Table 3. The results recorded in the ‘BoD’ column were obtained using the BoD super-efficiency model described by equation (2). As in the previous case, the values of the efficiency score and the ranks of universities are given. The next column, ‘BoD-R’, contains the results of the super-efficiency model described by equation (2), with additional weight restrictions derived from equation (3). The last five columns (O1-O5) contain the non-normalised values of all the sub-indicators.

Based on the results of the BoD model, Go8 group universities are at the top of the ranking, as illustrated in Table 3. UWA takes second place, UQ third place, MONASH fourth place, SYDNEY fifth place, MELBOURNE sixth place, ANU seventh place and ADELAIDE eighth place. First place is occupied by CDU (from the IRU group), mainly because it has the highest value for the O5 sub-indicator, indicating that it obtained the largest amount of grants. From the Go8 group, UNSW has a performance score of 0.861 and is ranked twelfth, which results from its sub-indicators being 20-40% lower than those of the leaders.

In the case of BoD-R model, the Go8 group universities are also at the top of the ranking, as illustrated in Table 3. MONASH takes first place, UQ second place, SYDNEY third place, UWA fourth place, MELBOURNE sixth place, ANU seventh place and ADELAIDE eighth place. Fifth place is occupied by CDU. From the Go8 group, UNSW is ranked eleventh with an efficiency score of 0.861, as in the previous model.

Table 4. Comparison of the results for the two BoD-R and BoD models for selected universities

University	BoD-R		BoD		Non-zero weights in BoD model				
	Score	Rank	Score	Rank	O1	O2	O3	O4	O5
CQU	0.535	34	0.698	23		X			
CDU	1.053	5	1.197	1		X			
MURDOCH	0.587	29	0.594	34	X	X	X		
RMIT	0.618	26	0.634	31	X	X			
SCU	0.630	25	0.640	30	X	X	X		
SWINBURNE	0.642	24	0.658	28	X	X			
FEDUNI	0.506	36	0.642	29		X			
USC	0.586	31	0.701	22		X			
VU	0.614	27	0.709	21		X			

The rankings for some universities vary considerably due to the varied weighting schemes. In the case of CI, the weights are fixed for all the sub-indicators. BoD provides for full flexibility in determining the weights, which allows for the assigning of zero weights to some sub-indicators, whereas BoD-R introduces weight restrictions, which eliminates the problem of zero weights; therefore, all the sub-indicators are considered when calculating the efficiency score. Table 4 presents selected results illustrating the differences in the positions of individual universities in the rankings based on the BoD and BoD-R methods. The analysis covers the universities with positions in the two rankings differing by at least 4 ranks (deterioration or improvement). Among the remaining universities, 14 did not change their position, 4 moved 1 place, 4 moved 2 places, and 6 moved 3 places.

The ‘Score’ columns show the values of the efficiency scores, and the ‘Rank’ columns show the universities’ places in the rankings for the BoD-R and BoD models. The columns labelled O1-O5 indicate the variables for which the BoD model has non-zero weights. It appears that in as many as five cases, the efficiency scores are calculated on the basis of only one sub-indicator.

Four of the universities listed in the above table improved their positions in the ranking: MURDOCH, RMIT and SCU by 5 places and SWINBURNE by 4 places. Five universities fell in the ranking: CQU by 11 places, CDU by 4 places, FEDUNI by 7 places, USC by 9 places and VU by 6 places. This outcome was the result of the influence of the zero weights in the BoD method, thus allowing the full flexibility of weighting. The values of the efficiency scores in the BoD-R method considered all of the variables: both those with a positive effect on efficiency and those that, for a given university, have inferior values compared to the other universities. These two methods show how much zero weights affect the overestimation or underestimation of efficiency scores.

Despite some differences among the three rankings, they are highly positively correlated. The correlation coefficient is 0.89 for the relation between the CI and BoD rankings, 0.95 for the relation between CI and BoD-R and 0.97 for the relation between BoD and BoD-R. Such high correlations indicate high convergence of results, which is confirmed by the high positions of the research universities from the Go8. This confirms the correctness of the results obtained and indicates that the problem of zero weights in this type of model must be solved by introducing additional restrictions on weights. The results from the BoD-R model are used for further considerations.

From a practical point of view, a very useful benefit of the BoD models based on DEA is their ability to calculate projections of variable changes, which would enable inefficient universities to achieve full efficiency. This ability is illustrated in Table 5, in which the projected non-normalised values needed to achieve full efficiency are calculated based on the BoD-R model for all the inefficient universities.

Table 5. Projection of sub-indicator values for all inefficient universities—BoD-R model

University	Score	O1	O2	O3	O4	O5
ADELAIDE	0.933	0.465	0.514	3.27	11.60	12.60
UTAS	0.907	0.376	0.571	3.56	11.96	12.37
CURTIN	0.880	0.465	0.513	3.30	11.10	12.43
UNSW	0.861	0.420	0.534	3.27	11.04	12.72
MACQUAIRIE	0.833	0.566	0.382	3.30	11.01	13.20
LATROBE	0.823	0.532	0.519	2.82	9.59	12.00
NEWCASTLE	0.817	0.501	0.520	3.02	10.23	12.16
UNISA	0.786	0.486	0.535	3.04	10.31	12.09
JCU	0.774	0.414	0.560	3.39	11.43	12.23
FLINDERS	0.747	0.422	0.558	3.36	11.31	12.22
UOW	0.732	0.514	0.493	3.08	10.39	12.37
QUT	0.717	0.532	0.519	2.82	9.59	12.00
DEAKIN	0.710	0.452	0.538	3.25	10.98	12.25
UTS	0.687	0.439	0.545	3.31	11.16	12.26
GRIFFITH	0.678	0.411	0.497	3.72	13.32	13.17
ECU	0.651	0.532	0.519	2.82	9.59	12.00
SWINBURNE	0.642	0.534	0.511	2.85	9.68	12.08
SCU	0.630	0.517	0.420	3.42	11.67	13.16
RMIT	0.618	0.543	0.473	2.98	10.07	12.41
VU	0.614	0.532	0.519	2.82	9.59	12.00
ACU	0.589	0.377	0.573	3.58	12.02	12.31
MURDOCH	0.587	0.533	0.422	3.32	11.09	12.97
UNE	0.586	0.547	0.459	3.03	10.21	12.52
USC	0.586	0.388	0.569	3.52	11.84	12.29
UWS	0.575	0.389	0.569	3.52	11.83	12.28
CANBERRA	0.552	0.493	0.459	3.39	11.36	12.82
CQU	0.535	0.377	0.573	3.58	12.02	12.31
CSU	0.530	0.537	0.498	2.90	9.82	12.19
FEDUNI	0.506	0.414	0.561	3.40	11.44	12.24
USQ	0.505	0.532	0.519	2.82	9.59	12.00

The smallest changes needed at these universities are for the O2 sub-indicator (64% on average)—the number of graduates of science disciplines—and the O1 sub-indicator (90% on average)—the number of ‘doctorates by research’—and these results correspond to the clear teaching orientation of these universities. This orientation towards teaching affects the universities’ small number of publications and citations and the small amount of awarded grants. For these universities to achieve full efficiency equal to that of the leaders, on average, their number of publications should increase by 2.5 times, their number of citations should increase by 4.0 times and the amount of grants they receive should increase by 3.6 times. Of course, this increase would make sense only if these universities wanted to compete in the field of scientific research.

Concluding remarks

The results obtained in this study confirm the usefulness of the methods applied to evaluate the research activities of universities. Two methods of measurement and evaluation are presented based on objective statistical data derived from reliable sources: the Australian Department of Education and Training and WoS. The methods are based on two approaches—CIs and BoD in two variants. The empirical verification of these methods using data from 2015 at 37 public Australian universities showed varied results because of the imperfections of the methods. The adopted weighting scheme was found to be a fundamental problem causing these discrepancies, as illustrated by the results described in the previous section. This problem was solved by supplementing the basic BoD algorithm with additional constraint conditions, eliminating the problem of zero weights and ensuring that all variables are considered when calculating the efficiency score. This result also confirms the need to use several computational methods in analyses of this type.

The model proposed in this article uses data regarding the numbers of publications and citations from WoS, despite other Australian authors using an Australia-specific weighted publication rate without considering citations. This procedure will allow future research into international comparisons of the Australian higher education system.

It is worth emphasising the practical utility of the empirically proven methods. The obtained results, in addition to building a ranking, also allow for broader analyses of the causes of a better or worse position in the ranking. If CIs are used, it is possible to indicate the strengths and weaknesses of each university. In the case of the BoD model, it is possible to calculate projections for worse performing universities. The projection allows the quantification of the directions of desired changes aimed at improving the performance of the research activities of these universities. It is equally important to rank fully efficient universities, which enables placing them in sequence. All of this information can be used both for operational management at the university level and for implementing research development policy across the entire higher education system.

Some universities are clearly research oriented, as evidenced by their participation in such groups as the Go8, IRU or ATN. The results of the study presented in this article indicate that these universities occupy top positions in the discussed rankings (especially as far as the Go8 group is concerned), confirming the reliability of such results.

It is obvious that the activities of universities comprise not only research but also teaching. Consequently, further research on Australian universities conducted by the author will focus on the measurement and evaluation of their teaching activity and the mutual relationships between research and teaching. It is well known that ensuring an appropriate level of teaching requires high competencies among academic staff, which are acquired by conducting intensive scientific research.

References

Abbott, M., & Doucouliagos, C. (2003). The efficiency of Australian universities: A data envelopment analysis. *Economics of Education Review*, *22*, 89-97. https://doi:10.1016/S0272-7757(01)00068-1.

Abbott, M., & Doucouliagos, H. (2004). Research output of Australian universities. *Education Economics*, *12*, 251-265. https://doi:10.1080/0964529042000258608.

Agasisti, T., Dal Bianco, A., Landoni, P., Sala, A., & Salerno M. (2011). Evaluating the efficiency of research in academic departments: an empirical analysis in an Italian Region. *Higher Education Quarterly, 65*, 267-289. https:// doi:10.1111/j.1468-2273.2011.00489.x.

Agasisti, T., & Haelermans, C. (2016). Comparing efficiency of public universities among European countries: Different incentives lead to different performances. *Higher Education Quarterly, 70*, 81-104. https://doi:10.1111/hequ.12066.

Allen, R., Athanassopoulos, A., Dyson, R.G., & Thanassoulis E. (1997). Weights restrictions and value judgements in data envelopment analysis: Evolution, development and future directions. *Annals of Operations Research*, *73*, 13-34. https://doi.org/10.1023/A:1018968909638.

Andersen, P., & Petersen, N. C. (1993). A procedure for ranking efficient units in data envelopment analysis. *Management Science*, *39*, 1261-1264. https://doi:10.1287/mnsc.39.10.1261.

Angulo-Meza, L., & Lins, M. P.E. (2002). Review of methods for increasing discrimination in data envelopment analysis. *Annals of Operations Research*, 116(1-4), 225-242. https://doi.org/10.1023/A:1021340616758.

Avkiran, N. K. (2001). Investigating technical and scale efficiencies of Australian universities through data envelopment analysis. *Socio-Economic Planning Sciences*, *35*, 57-80. https://doi:10.1016/S0038-0121(00)00010-0.

Barra, C., & Zotti, R. (2016). Measuring efficiency in higher education: An empirical study using a bootstrapped data envelopment analysis. *International Advances in Economic Research*, *22*, 11-33. https://doi:10.1007/s11294-015-9558-4.

Beasley, J. (1990). Comparing university departments. *Omega – The International Journal of Management Science*, *18*, 171-183. https://doi.org/10.1016/0305-0483(90)90064-G.

Carrington, R., Coelli, T., & Rao, D. S. P. (2005). The performance of Australian universities: Conceptual issues and preliminary results. *Economic Papers: A Journal of Applied Economics and Policy*, *24*, 145-163. https://doi:10.1111/j.1759-3441.2005.tb01001.x.

Carrington, R., O'Donnell, C., & Rao, D. S. P. (2016). Australian university productivity growth and public funding revisited. *Studies in Higher Education*, 1–22. https://doi:10.1080/03075079.2016.1259306.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*, 429-444. https://doi:10.1016/0377-2217(78)90138-8.

Cherchye, L., Moesen, W., Rogge, N., & Van Puyenbroeck, T. V. (2007). An introduction to 'benefit of the doubt' composite indicators. *Social Indicators Research*, *82*, 111–145. https://doi:10.1007/s11205-006-9029-7.

Cherchye, L., Moesen, W., Rogge, N., & Van Puyenbroeck, T. (2011). Constructing composite indicators with imprecise data: A proposal. *Expert Systems with Applications*, *38*, 10940-10949. https://doi:10.1016/j.eswa.2011.02.136.

Cook, W. D., Tone, K., & Zhu, J. (2014). Data envelopment analysis: Prior to choosing a model. *Omega-The International Journal of Management Science*, *44*, 1-4. https://doi:10.1016/j.omega.2013.09.004.

Cooper, W. W., Seiford, L. M., & Zhu, J. (2011). *Handbook on data envelopment analysis*. New York, NY: Springer.

Cooper, W.W., Ruiz, J. L., & Sirvent, I. (2009). Selecting non-zero weights to evaluate effectiveness of basketball players with DEA. *European Journal of Operational Research*, *195*, 563-574. https://doi:10.1016/j.ejor.2008.02.012.

De Witte, K., & Rogge, N. (2010). To publish or not to publish? On the aggregation and drivers of research performance. *Scientometrics*, *85*, 657-680. https://doi:10.1007/s11192-010-0286-5.

Despotis, D. K. (2005). Measuring human development via data envelopment analysis: The case of Asia and the Pacific. *Omega – The International Journal of Management Science*, *33*, 385-390. https://doi.org/10.1016/j.omega.2004.07.002.

DET. (2017). Australian Government. Department of Education and Training. Retrieved from https://www.education.gov.au.

Docampo, D., & Cram, L. (2017). Academic performance and institutional resources: A cross-country analysis of research universities. *Scientometrics*, *110*, 739-764. https://doi:10.1007/s11192-016-2189-6.

Filippetti, A., & Peyrache, A. (2011). The patterns of technological capabilities of countries: A dual approach using composite indicators and data envelopment analysis. *World Development*, *39*, 1108-1121. https://doi:10.1016/j.worlddev.2010.12.009.

Freudenberg, M. (2003). Composite indicators of country performance: A critical assessment. *OECD Science, Technology and Industry Working Papers, 2003/16*. Paris: OECD Publishing. http://dx.doi.org/10.1787/405566708255.

Gnaldi, M., & Ranalli, M. G. (2016). Measuring university performance by means of Composite Indicators: A robustness analysis of the composite measure used for the benchmark of Italian universities. *Social Indicators Research*, *129*, 659-675. https://doi:10.1007/s11205-015-1116-1.

Johnes, J. (2006). Data envelopment analysis and its application to the measurement of efficiency in higher education. *Economics of Education Review*, *25*, 273-288. https://doi:10.1016/j.econedurev.2005.02.005.

Katharaki, M., & Katharakis, G. (2010). A comparative assessment of Greek universities' efficiency using quantitative analysis. *International Journal of Educational Research*, *49*, 115-128. https://doi:10.1016/j.ijer.2010.11.001.

Koopmans, T. C. (1951). Analysis of production as an efficient combination of activities. In T. C. Koopmans (Ed.), *Activity Analysis of production and allocation* (pp. 33-97). New York: Wiley.

Lee, B. L. (2011). Efficiency of research performance of Australian universities: A reappraisal using a bootstrap truncated regression approach. *Economic Analysis and Policy*, *41*, 195-203. https://doi:10.1016/S0313-5926(11)50032-3.

Lovell, C. A. K., Pastor, J. T., & Turner, J. A. (1995). Measuring macroeconomic performance in the OECD: A comparison of European and non-European countries. *European Journal of Operational Research*, *87*, 507-518. https://doi:10.1016/0377-2217(95)00226-X.

Lukman, R., Krajnc, D., & Glavič, P. (2010). University ranking using research, educational and environmental indicators. *Journal of Cleaner Production*, *18*, 619-628. https://doi:10.1016/j.jclepro.2009.09.015.

Malul, M., Hadad, Y., & Ben-Yair, A. (2009). Measuring and ranking of economic, environmental and social efficiency of countries. *International Journal of Social Economics*, *36*, 832-843. https://doi:10.1108/03068290910967109.

Manca, A. R., Governatori, M., & Mascherini, M. (2010). *Towards a set of composite indicators on Flexicurity: A comprehensive approach*. Luxembourg, Belgium: European Commission, Joint Research Centre, European Communities.

Marconi, G., & Ritzen, J. (2015). Determinants of international university rankings scores. *Applied Economics*, *47*, 6211-6227. https://doi:10.1080/00036846.2015.1068921.

Marginson, S. (2007). Global university rankings: Implications in general and for Australia. *Journal of Higher Education Policy and Management*, *29*, 131-142. https://doi:10.1080/13600800701351660.

Marginson, S. (2014). University rankings and social science. *European Journal of Education*, *49*, 45-59. https://doi:10.1111/ejed.12061.

Mecit, E. D., & Alp, I. (2013). A new proposed model of restricted data envelopment analysis by correlation coefficients. *Applied Mathematical Modelling*, 37, 3407-3425. http://dx.doi.org/10.1016/j.apm.2012.07.010.

Moesen, W., & Cherchye, L. (1998). The macroeconomic performance of nations. Measurement and perception. Discussion paper Series 98.22. Leuven, Belgium: Centre for Economic Studies.

Murias, P., de Miguel, J. C., & Rodríguez, D. (2008). A composite indicator for university quality assessment: The case of Spanish higher education system. *Social Indicators Research*, *89*, 129-146. https://doi:10.1007/s11205-007-9226-z.

Organization for Economic Co-Operation and Development. (2008). *Handbook on constructing composite indicators. Methodology and user guide*. Paris: OECD Publications.

Paruolo, P., Saisana, M., & Saltelli, A. (2013). Ratings and rankings: Voodoo or science? *Journal of the Royal Statistical Society: Series A*, *176*, 609-634. https://doi:10.1111/j.1467-985X.2012.01059.x.

Pastor, J. M., Serrano, L., & Zaera, I. (2015). The research output of European higher education institutions. *Scientometrics*, *102*, 1867-1893. https://doi:10.1007/s11192-014-1509-y.

Podinovski, V. V., & Thanassoulis, E. (2007). Improving discrimination in data envelopment analysis: some practical suggestions. *Journal of Productivity Analysis*, *28*, 117-126. https://doi:10.1007/s11123-007-0042-x.

Rogge, N. (2011). Granting teachers the "benefit of the doubt" in performance evaluations. *International Journal of Educational Management*, *25*, 590-614. https://doi:10.1108/09513541111159077.

Sagarra, M., Mar-Molinero, C., & Agasisti T. (2017). Exploring the efficiency of Mexican universities: Integrating data envelopment analysis and multidimensional scaling. *Omega – The International Journal of Management Science, 67*, 123–133. https://doi.org/10.1016/j.omega.2016.04.006.

Sarrico, C. S., & Dyson, R. G. (2004). Restricting virtual weights in data envelopment analysis. *European Journal of Operational Research*, *159*, 17-34. https://doi:10.1016/S0377-2217(03)00402-8.

Shen, Y., Ruan, D., Hermans, E., Brijs, T., Wets, G., & Vanhoof, K. (2011). Modeling qualitative data in data envelopment analysis for composite indicators. *International Journal of System Assurance Engineering and Management*, *2*, 21-30. https://doi:10.1007/s13198-011-0051-z.

The Australian Research Council. (2015). *State of Australian university research 2015–16*. Accessed October 2016. Retrieved from http://www.arc.gov.au/era-reports.

Valadkhani, A., & Worthington, A. (2006). Ranking and clustering Australian university research performance, 1998–2002. *Journal of Higher Education Policy and Management*, *28*, 189-210. https://doi:10.1080/13600800600751101.

Van Puyenbroeck, T. (2017). On the output orientation of the benefit-of-the-doubt-model. *Social Indicators Research*. https://doi:10.1007/s11205-017-1734-x.

Wong, Y. H. B., & Beasley, J. E. (1990). Restricting weight flexibility in data envelopment analysis. *The Journal of the Operational Research Society*, *41*, 829-835. https://doi:10.2307/2583498.

WoS. (2017). *Web of science*. Retrieved from https://webofknowledge.com.

Worthington, A. C., & Lee, B. L. (2008). Efficiency, technology and productivity change in Australian universities, 1998–2003. *Economics of Education Review*, *27*, 285–298. https://doi:10.1016/j.econedurev.2006.09.012.

Zanella, A., Camanho, A. S., & Dias, T. G. (2015). Undesirable outputs and weighting schemes in composite indicators based on data envelopment analysis. *European Journal of Operational Research*, *245*, 517-530. https://doi:10.1016/j.ejor.2015.03.036.