A network DEA quantity and quality-orientated production model: An application to Australian university research services

Boon L. Lee [a], Andrew C. Worthington [b],*

[a] School of Economics and Finance, Queensland University of Technology, 2 George St, Brisbane, Queensland 4000, Australia
[b] Department of Accounting, Finance and Economics, Griffith University, 170 Kessels Rd, Nathan, Queensland 4111, Australia

ARTICLE INFO

ABSTRACT

The motivation for this analysis is the recently developed Excellence in Research for Australia (ERA) program developed to assess the quality of research in Australia. The objective is to develop an appropriate empirical model that better represents the underlying production of higher education research. In general, past studies on university research performance have used standard DEA models with some quantifiable research outputs. However, these suffer from the twin maladies of an inappropriate production specification and a lack of consideration of the quality of output. By including the qualitative attributes of peer-reviewed journals, we develop a procedure that captures both quality and quantity, and apply it using a network DEA model. Our main finding is that standard DEA models tend to overstate the research efficiency of most Australian universities.

1. Introduction

The motivation for this study is the Australian Commonwealth government's Excellence in Research for Australia (ERA) program, recently developed to assess the quality of research in Australian universities. Implemented in 2008, the ERA initiative was a key element in the previous Labor government's agenda for the reform of Australia's higher education system as a means of providing assurances to government, industry, business, and others stakeholders of the quality of research conducted in Australian universities. Moreover, the ERA was aimed not only at improving the overall level of research quality in Australia, but also '…sits within a broader movement internationally, with the emphasis on excellence and quality being found in the performance-based funding schemes of a number of other countries' [1].

There is then clearly a need for an appropriate measure of the productive efficiency of research. This is not least because of the need to account for the billions of dollars of research income (both public and private) the sector attracts, the significant contribution this research makes to the Australian economy, and for the purpose of enhancing the domestic and international reputations of both individual institutions and the sector as a whole. In doing so, it should create greater transparency and accountability of publicly funded institutions and appropriately recognize

the efforts of those universities producing research and therefore deserving of funding. As Herbst [2] points out, 'The rationale of performance funding is that funds should flow to institutions where performance is manifest: 'performing' institutions should receive more income than lesser performing institutions, which would provide performers with a competitive edge and would stimulate less performing institutions to perform. Output should be rewarded, not input'.

To meet this need, we set two objectives for this analysis. First, the development of a suitable production model that best represents university research activity while considering the appropriate inputs that generate outputs. One of the most contentious issues in existing higher education efficiency studies is that the number of research publications produced and the amount of grants received are both considered outputs, yet may also comprise inputs [3–9]. Without exception, these studies employ standard data envelopment analysis (DEA) to measure technical efficiency based on the production assumption of some set of inputs used to generate some set of outputs. However, the rigidity of standard DEA treats the production process as a 'black box' in that it simply transforms inputs into outputs and neglects any possible intervening processes, including dissimilar series or parallel functions.

In the case of the abovementioned studies, this has meant that while the number of publications and/or the value of grants are rightly outputs in their own right, it fails to illustrate the relationship between the variables and the mapping of the underlying production process. In reality, decision-making units (DMUs), here universities, often perform several different functions and readily

* Corresponding author. Tel.: +61 7 3735 4273; fax: +61 7 3735 3719.
 E-mail addresses: bl.lee@qut.edu.au (B.L. Lee),
a.worthington@griffith.edu.au (A.C. Worthington).

separate these functions into different components, in a series or in parallel and/or in a more complex form of network type. This suggests that the outputs produced in a certain series may become intermediate inputs in a subsequent production stage. To determine the appropriate production process that best represents university research production, we employ the network DEA (NDEA) approach.

The second objective of our analysis relates to a rather more practical concern regarding the measurement of research output in the Australian context. The ERA 2010 surveyed all Australian universities on their journal publications over the period 2003–08, categorizing each journal into one of four broad groupings. These were A* (virtually all papers are of very high quality), A (the majority of papers are of very high quality), B (a few papers are of very high quality) and C journals (the papers are peer reviewed and of some quality, but do not meet the criteria of the higher tiers). Subsequently, ERA 2012 evaluated research undertaken between 1 January 2005 and 31 December 2010. However, unlike ERA 2010, ERA 2012 did not provide any journal categorization other than meeting the minimum requirement of a peer-reviewed publication. The lack of a quality indicator thus questions whether we accurately measure the research performance of Australian universities. It also raises questions on how funding is allocated where there is no attempt to identify quality research publications.

In addition, the lack of any indicator of quality leaves us uncertain as to whether the standard of university research publication has improved throughout the (albeit limited) ERA exercise to date. To better measure the quality of research publication, we propose an approach that aggregates the number of publications for each university using weights. We detail the procedure in the data section. An essential aspect in deriving a qualitative dataset of research publications is that based on our production model, it is an intermediate measure, which could influence efficiency scores in its various stages. This suggests that the lack of a reliable research publication dataset could distort efficiency scores, especially in the second stage of our production model where both the quantity and quality of research outputs determine the allocation of research grants to universities.

The remainder of the paper is structured as follows. Section 2 describes the basic NDEA framework used for our research production model. Section 3 details the data sources, including the methodology for measuring research publication quality and quantity. Section 4 presents the empirical results. Section 5 provides some brief remarks.

2. Conceptual framework for the research production model

As elsewhere around the world, the research production model in Australian universities is complex. Perusing the many higher education studies employing standard DEA, we can see that it is common to specify the inputs as full-time equivalent (FTE) academic staff and capital stock (or some proxy) and the outputs as the number of publications and the value of grants. The problem with this model is that the inputs used to produce outputs may not be rational because of incorrect sequencing. There is then a need to adopt an approach that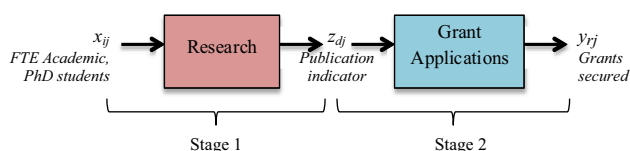 takes into account a network of 'divisions' or 'sub-processes' to appropriately assess the divisional efficiencies and the overall efficiency of universities.

One such approach is the NDEA model [10–12], which has been employed in numerous studies across a wide range of industries, including airports [13–14], banks [15–20], hotels [21], electric utilities [22–24], university libraries [25] and research and development [26–27]. In brief, NDEA is made up of a network of 'divisions' or 'nodes', which when viewed together, comprise the individual DMU. Each node transforms inputs into outputs, and in some nodes, the outputs become inputs themselves to produce yet other outputs in yet other nodes. Each node generates an efficiency (i.e. divisional efficiency) score that we can compare with the corresponding node of other DMUs. We evaluate overall efficiency score using the final node. Importantly, depending on the precise nature of the nodes, we may attach different weights to each node to recognize the particular emphasis of the production model.

Fig. 1 depicts our NDEA model for universities. The model comprises two stages. In the first stage, inputs x_{ij} produce intermediate outputs z_{dj}. In our university research production model, we use two primary inputs in Stage 1 (Node 1), namely, FTE academics and doctorates by research as student load for the j-th university. We consider only PhD students because of their longer period of candidature, suggesting that they more readily engage in research collaboration and co-authorship with academics than say, coursework only or shorter-period masters research students. This is especially the case where PhD students quite often serve as research assistants in the production of publications, while it is increasingly common for PhD students in Australia to submit their theses by publication or as a series of published papers, thus evidencing that PhD students are an input in publication [28].

We specify a single output for Stage 1, which is the publication indicator described later. While we acknowledge that PhD completions are also an output in the first stage, our production model only focuses on the outcomes of research activity. In the second stage, z_{dj} becomes an intermediate input in the process of grant application in order to win grants (i.e. grants secured) denoted by y_{rj}. Hence, in Stage 2 (Node 2), the publication indicator becomes an intermediate input and grants awarded our final output.

3. Data sources and method

3.1. Data

The data consists of 37 annual observations of Australian universities over the period 2004–11. All data are from the Higher Education Statistics and Financial Reports compiled by the Commonwealth Department of Education (www.education.gov.au) [previously Department of Education, Science, and Training (DEST) and Department of Education, Employment and Workplace Relations (DEEWR)]. We specify a production model where the research production process includes only research-relevant inputs. The primary inputs are FTE academics (comprising 'research-only' academic staff and 'teaching and research' academic staff), and doctorates by research (student load). As the intermediate output, we consider the publication indicator. The final output for the NDEA is research income (i.e. grants won), comprising research income from across the designated categories of Australian Competitive Grants, Other Public Sector Grants, Industry and Other Grants, and Cooperative Research Centers (CRC).

The measurement of university performance is critically dependent on the reliability of the research publication output, which should incorporate both quality and quantity. There are many

Fig. 1. A university research production model.

existing studies on university performance from round the world, but most face the significant obstacle of incorporating quality attributes in research publications. As a result, they either omit quality attributes altogether or consider a proxy for quality, which may not be fully appropriate. For example, in the Australian context, [29–31] measured the research output of Australian universities using, among other things, the number of publications (including books, book chapters, journal articles, and refereed conference papers), but did not account for any quality attributes. Elsewhere, [32] measured the research productivity of universities in the US using total publications and the ratio of publications to full-time faculty, but the approach was nonetheless essentially quantitative and excluded any specific qualitative measure.

In a similar fashion, [33] used the number of publications in Chinese higher education and the reputations of and publications per academic staff as outputs to reflect the characteristics of research, while [34] distinguished between the number of overseas and domestic publications by Chinese universities perhaps to capture quality differences, although this was not specifically stated. Elsewhere, [35] went a step further by categorizing research publications into five groups, of which one comprised core economic journals. However, putting aside that their study focused solely on the efficiency of economics departments, there was still no clear delineation of quality differences across journals. Fundamentally, the major limitation in using the number of publications as an output is that it does not capture the quality aspect of research outputs for journals as no weights attach to higher-ranked journals. Moreover, even within top-tier journals, quality differences exist and different journal practices are common across disciplines [5].

Partly in response, other studies have specified external research funding as a quality-adjusted measure of output, including [2,36,7,37–41] This is because it not only provides a market valuation of research, but also according to [6]), represents some overall external perception of quality. However, this is an aggregated measure and does not tell us anything about the quality of the research publications themselves. Furthermore, there may be distortions associated with the differences in research costs across academic disciplines [6], e.g. low-cost social sciences vs. high-cost physical sciences. Finally, as discussed, a remaining problem with

Table 1
Descriptive statistics.

Input/output	Mean	Std. dev.	Min.	Max.
2004				
FTE academic staff	917	688	101	2533
Doctorate by research (Student load)	732	649	33	2481
Publication indicator (unweighted)	806	811	44	2986
Publication indicator (weighted)	973	1210	24	4226
Total research income ($)	40,036,418	50,458,436	695,047	198,739,216
2005				
FTE academic staff	952	729	101	2679
Doctorate by research (Student load)	753	663	38	2533
Publication indicator (unweighted)	677	661	36	2650
Publication indicator (weighted)	1384	1687	57	6309
Total research income ($)	43,316,797	53,394,818	757,969	208,281,603
2006				
FTE academic staff	974	744	121	2781
Doctorate by research (Student load)	771	679	40	2537
Publication indicator (unweighted)	585	578	25	2110
Publication indicator (weighted)	1476	1804	54	6412
Total research income ($)	49,345,611	62,079,097	731,364	254,441,240
2007				
FTE academic staff	1002	781	144	2858
Doctorate by research (Student load)	787	696	60	2575
Publication indicator (unweighted)	994	952	103	3475
Publication indicator (weighted)	1428	1766	73	6642
Total research income ($)	59,576,513	78,610,819	1,023,803	309,488,777
2008				
FTE academic staff	1029	801	173	2931
Doctorate by research (Student load)	809	712	70	2615
Publication indicator (unweighted)	1071	982	143	3663
Publication indicator (weighted)	1612	1960	113	7573
Total research income ($)	67,266,149	87,630,378	2,049,559	310,660,056
2009				
FTE academic staff	1060	835	187	3055
Doctorate by research (Student load)	854	741	84	2670
Publication indicator (unweighted)	1146	1062	139	3848
Publication indicator (weighted)	1819	2184	113	8167
Total research income ($)	75,791,752	100,985,641	2,377,018	382,546,727
2010				
FTE academic staff	1077	832	210	3071
Doctorate by research (Student load)	921	786	101	2831
Publication indicator (unweighted)	1176	1091	135	4082
Publication indicator (weighted)	1918	2303	123	8746
Total research income ($)	75,000,195	94,561,321	3,058,290	336,973,307
2011				
FTE academic staff	1088	840	193	3181
Doctorate by research (Student load)	972	816	105	2924
Publication indicator (unweighted)	1242	1141	140	4285
Publication indicator (weighted)	2257	2690	147	10,346
Total research income ($)	87,685,896	109,066,183	2,993,628	376,531,487

funding as a qualitative measure is that it is an input in the research process [3,42]. Thus, the use of research funding as a proxy for research output is not a suitable measure to reflect the quality aspect of research publication.

Another quality measure, which most academics generally accept as a quality indicator, is the number of citations. For example, [43] specifies the number of citations from the ISI-Web of Science database. While their research output measure incorporates both quality and quantity, there are drawbacks in relying solely on the number of citations as a qualitative output. For instance, [28] noted that the suggestion that citations necessarily represent some useful contribution to knowledge and society might not always be the case. For example, citations may include articles that attract negative criticism rather than positive acknowledgment and recognition of their contribution toward the literature and stakeholder's welfare. The problem of citation bias also arises from the practice of self-citation.

As alternatives, [44] argue that peer reviews are a better quality measure than citations, which strengthens our approach to measuring quality research output. [45] also argues that the drawback in number of citations is the variation in shelf life across different disciplines. Yet another quality aspect of a journal is its reputation, which is rather subjective, as in [46]. In addition, citations are only useful at the aggregate level when comparing individual journal articles or academics. At the disaggregated level, such as on an annual basis, this becomes irrelevant as the citation of past studies only reflects current performance and not the year published. Furthermore, articles published at the start of a year have a greater opportunity of citation than those published at the end of the year during the reporting period, which can introduce a (small) bias in performance. Table 1 presents statistics for the inputs and outputs used in the NDEA production model.

3.2. Development of journal dataset

Our weighted publications dataset comprises authored books, book chapters, and journal articles. Given the lack of any comprehensive information concerning the quality of books and book chapters, we assume the same weights used in the Higher Education Statistics where authored books carry a weight of 5.00 and book chapters a weight of 1.00. For the journal publications, we use data from Scopus and the SCImago Journal Rank (SJR) indicator list to create the appropriate weight. Since 2000, Scopus has been the world's largest scientific database, covering most of the journals included in the Thomson-Reuters' Web of Science and other databases [47]. This makes it an appropriate database when aggregating the total number of publications by university. The SJR database (www.scimagojr.com) covers over a period of fourteen years (1999–2013). In 2003, it included some 17,512 academic journals, increasing its coverage to 29,385 journals in 2013.

The SJR is a size-independent metric, which measures the current 'average prestige per paper' of journals [47]. Thus, changes in the prestige or influence and impact factor of the journal determine the SJR indicator for a particular journal, which varies over time. The primary purpose of the SJR is journal evaluation. Recent studies using the SJR include [48], who used it as a tool for evaluating journals in the Scopus database, [49], who compared the SJR with the Thomson Scientific impact factor, and [50], who grouped the impact factor and the SJR together to examine both bibliometric and usage indicators.

It is worth noting that one tradeoff of producing quality journal articles is the loss of the additional number of articles possible in lesser quality journals. Hence, we note the tradeoff between quality and quantity. If the measurement of output only includes quantity, then universities that focus mainly on publishing in high quality journals will suffer a penalty because of the lack of contribution from a quantitative standpoint. By incorporating qualitative attributes into the measurement of output, we compensate for this. Hence, it is important that the research output measures include quality attributes.

Incorporation of quality in our aggregation requires a few steps. To start, we obtain the SJR indicator list from the official website for a specific year. Next, in SCOPUS we select a university and refine the search by document and source type. Refining the search provides a list of journals that the university has had papers published in. Let us call this the SCOPUS list. Then using a macro in MS Excel, we use the value attached to each journal from the SJR indicator list, and assign that value to the SCOPUS list. Table 2 shows a small sample of the SCOPUS list for a university and the SJR indicator list. When we retrieve the SCOPUS list, we copy it into MS Excel, as shown in Table 2. We obtain the assigned value next to it using the macro 'vlookup', which attempts to match the SCOPUS list with the SJR journal title list. Note that there are literally thousands of journal titles under the SJR indicator list, which makes it impossible to provide all in one table, and thus we only provide a sample.

This assigning of a value to each journal on the SCOPUS list essentially quality weights each journal. When all journals have been assigned a value, we aggregate the values on the SCOPUS list (for each university) to obtain the quality-adjusted publication indicator for each year for each university. From Table 2, this is the sum of $0.599+23.389+....+19.466=83.067$. As far as the authors are aware of, no existing study has attempted to develop a longitudinal database of university research publications using SJR and SCOPUS nor any other comparable approach.

3.3. Network DEA

We employ the centralized NDEA model in [51] because it satisfies the flow process form of Fig. 1. Following [37], we assume DMU_j ($j=1,2, …, n$) has D intermediate measures, z_{dj} ($d=1,2, …,D$).

Table 2
Illustration of assigning SJR value to SCOPUS list for a university.

SCOPUS list for university	Assigned value	Journal title	SJR indicator
Quality-adjusted publication indicator	83.067		
Social Psychology of Education	0.599	Annual Review of Immunology	36.731
Reviews of Modern Physics	23.389	Annual Review of Biochemistry	33.927
J of Early Christian Studies	0.124	Cell	27.227
Reviews of Modern Physics	23.389	Annual Rev of Cell and Develop Biology	23.581
J of Science and Medicine in Sport	0.398	Reviews of Modern Physics	23.389
Australian J of Educat and Develop Psych	0.103	Annual Review of Neuroscience	23.175
Quarterly J of Economics	15.599	Nature Reviews Molecular Cell Biology	20.325
Physiological Reviews	19.466	Physiological Reviews	19.466
		Genes and Development	17.883

Table 3
University ranking based on average efficiency scores, 2004–11.

	DEA[1]	Unweighted			Weighted		
		Overall	Stage 1	Stage 2	Overall	Stage 1	Stage 2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Australian Catholic University	35	37	28	37	37	34	33
Central Queensland University	17	30	27	28	26	32	19
Charles Darwin University	33	1	17	1	1	16	1
Charles Sturt University	36	32	15	33	29	20	31
Curtin University of Technology	29	14	29	14	15	26	8
Deakin University	10	25	22	27	23	22	25
Edith Cowan University	34	31	24	31	28	28	24
Griffith University	15	20	21	19	20	21	21
James Cook University	20	18	19	18	16	9	30
La Trobe University	27	26	16	29	27	18	32
Macquarie University	4	22	5	30	24	11	35
Monash University	3	8	12	10	9	7	16
Murdoch University	7	12	32	4	12	24	2
Queensland University of Technology	9	16	25	16	17	25	10
Royal Melbourne Institute of Technology	6	28	36	21	31	35	18
Southern Cross University	19	23	37	15	25	37	4
Swinburne University of Technology	2	27	35	22	32	14	36
Australian National University	5	11	10	12	11	2	28
Flinders University of South Australia	13	13	2	17	13	12	20
University of Adelaide	11	5	11	5	5	8	5
University of Melbourne	26	2	4	2	2	1	15
University of New England	23	19	6	26	19	17	23
University of New South Wales	28	6	3	9	6	5	12
University of Newcastle	16	9	8	11	7	10	7
University of Queensland	12	7	13	6	8	6	14
University of Sydney	8	3	1	8	3	3	17
University of Western Australia	22	4	9	3	4	4	11
University of Ballarat	25	24	31	23	22	30	13
University of Canberra	18	21	18	24	18	27	9
University of South Australia	21	15	33	13	14	33	3
University of Southern Queensland	31	34	26	32	33	29	27
University of Tasmania	24	10	23	7	10	13	6
University of Technology, Sydney	14	29	34	25	30	31	22
University of the Sunshine Coast	32	36	20	36	35	23	37
University of Western Sydney	1	33	7	35	34	19	34
University of Wollongong	30	17	14	20	21	15	26
Victoria University of Technology	37	35	30	34	36	36	29

For DMU_j, we denote the efficiency for the first stage as e_j^1 and the second as e_j^2. Based upon the DEA model of [52], we define

$$e_j^1 = \sum_{d=1}^{D} w_d z_{dj} / \sum_{i=1}^{m} v_i x_{ij} \quad \text{and} \quad e_j^2 = \sum_{r=1}^{s} u_r y_{rj} / \sum_{d=1}^{D} \tilde{w}_d z_{dj} \tag{1}$$

where v_i, w_d, $\tilde{w}_d$, and u_r are unknown nonnegative weights. Note

that w_d is set equal to $\tilde{w}_d$ in [37] as this is a characteristic of the centralized model. This means that $e_o^1 \times e_o^2 =$

$$\sum_{r=1}^{s} u_r y_{ro} / \sum_{i=1}^{m} v_i x_{io} \tag{2}$$

Hence, the centralized model can be expressed as

$$e_o^{centralized} = Max \ e_o^1 \times e_o^2 = \sum_{r=1}^{s} u_r y_{ro} / \sum_{i=1}^{m} v_i x_{io} \tag{3}$$

s.t.

$$e_j^1 \leq 1 \ and \ e_j^2 \leq 1 \ and \ w_d = \tilde{w}_d$$

Eq. (3) can then be converted into the following linear program

$$e_o^{centralized} = \sum_{r=1}^{s} u_r y_{ro} / \sum_{i=1}^{m} v_i x_{io}$$

s.t.

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{d=1}^{D} w_d z_{dj} \leq 0, \ j = 1, 2, ..., n,$$

$$\sum_{d=1}^{D} w_d z_{dj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0, \ j = 1, 2, ..., n,$$

$$\sum_{i=1}^{m} v_i x_{io} = 1, \tag{4}$$

$$w_d \geq 0, \ d = 1, 2, ..., D; \ v_i \geq 0, \ i = 1, 2, ..., m; \ u_r \geq 0, \ r = 1, 2, ..., s.$$

Eq. (4) yields the overall efficiency of the two-stage process. Assume that Eq. (4) provides a unique solution. Then the efficiencies for the first and second stages are obtained and expressed as

$$x_{io} = \sum_{d=1}^{D} w_d^* z_{do} e_o^{1, \ centralized} = \sum_{d=1}^{D} w_d^* z_{do} / \sum_{i=1}^{m} v_i^* x_{io}$$

$$= \sum_{d=1}^{D} w_d^* z_{do} \ and \ e_o^{2, \ centralized} = \sum_{r=1}^{s} u_r^* y_{ro} / \sum_{d=1}^{D} w_d^* z_{do} \tag{5}$$

If we denote the optimal value to Eq. (4) as $e_o^{centralized}$, then we have an efficiency decomposition of $e_o^{centralized} = e_o^{1, \ centralized} \times e_o^{2, \ centralized}$, as also proposed by [53].

4. Empirical results

Table 3 presents ranks the universities according to their efficiency scores. The results in Column 1 are from a standard DEA output-oriented model under constant returns to scale (CRS). Column 2 is the overall efficiency scores based on the NDEA approach, which is the product of Columns 3 and 4 using unweighted publications. Column 5 includes the overall efficiency scores based on Eq. (4), and their decomposition into Stage 1 (Column 6) and Stage 2 efficiency scores (Column 7) based on Eq. (5). The basis of the efficiency scores in Columns 2 through 4 is unweighted publications, meaning that the research publication uses the basic data in the Higher Education Statistics. The results show that universities tend to perform better under the standard DEA model than the NDEA model. We expect this is because the standard DEA model treats efficiency evaluation as a 'black box' and fails to identify inefficiencies within the internal processes. It thus provides little insight into the sources of inefficiency and the operational stages where inefficiency may arise.

In contrast, with the NDEA model, the overall efficiency score is the product of the efficiencies of all stages of the production model. In this case, if either of the two stages of the university production model is inefficient, it will affect the overall efficiency score. From our observed efficiency scores, none of the DMUs
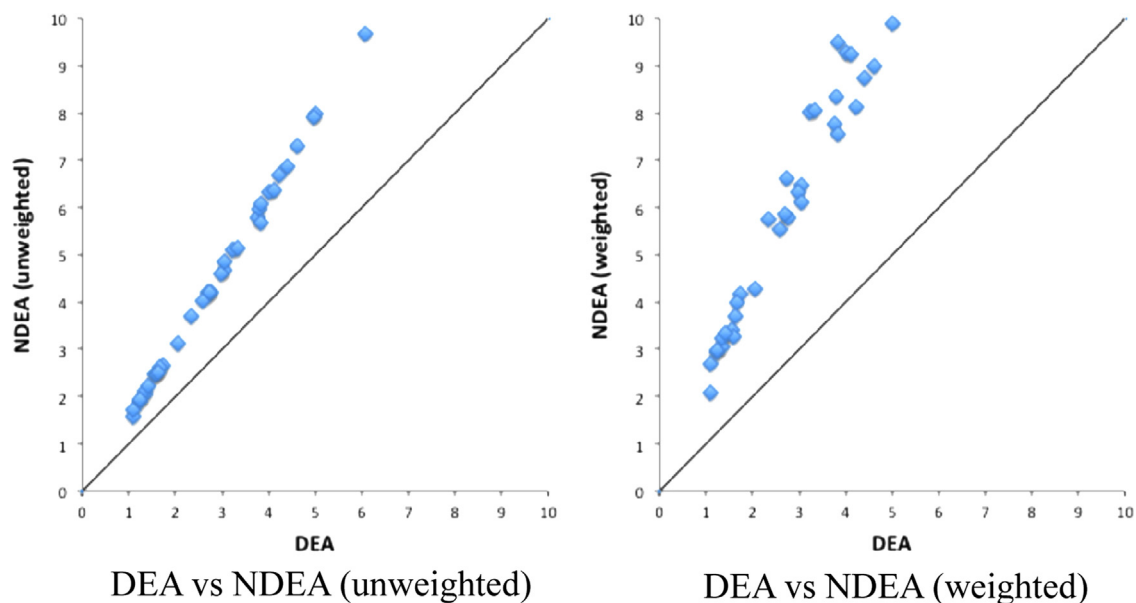
DEA vs NDEA (unweighted)   DEA vs NDEA (weighted)

Fig. 2. Efficiency scores for DEA vs. NDEA (unweighted) and DEA NDEA (weighted).

achieved an efficiency score of 1.00 (perfectly efficient) because of at least one stage having an efficiency score more than 1.00. Clearly, the inefficient scores are the result of an incorrect production model as it considered the inputs of FTE academic staff and actual load of doctoral students to produce grants. This flow-process failed to include the intermediate measure 'publications' and the history and quality of an applicant's publication, which is an essential input in being awarded grants.

When we compare the rankings of universities for unweighted publications to weighted publications, we observe that the Group of Eight (G8) leading universities have performed relatively well and rank in the top-ten in both Stage 1 and Stage 2. Overall, the G8 universities perform relatively better than, with the possible exception of Charles Darwin University.

From the analysis in Table 3, it suggests that using unweighted journal publications would exaggerate the performance of some universities that are not strong in research, but are still able to appear to perform well because what they offset quality with quantity. The ranking results in Table 3 therefore clearly illustrate the importance of using quality-weighted publications to assess the efficiency scores for university research performance.

Fig. 2 presents scatter plots of the efficiency scores between DEA vs NDEA (unweighted) and DEA vs NDEA (weighted). It is important to stress here that the efficiency scores in Fig. 2 are averages over the 2004–11 period, and because all universities have experienced at least some relative inefficiency over this period, we do not expect any universities to achieve perfect efficiency (i.e. 1.00) on average. The first observation from Fig. 2 is that NDEA scores are generally higher than the DEA scores, which is largely because of inefficiencies in either Stage 1 or 2 of the NDEA. The second observation is that the DEA vs NDEA (unweighted) comparisons indicate a linear relationship suggesting the lack of discriminatory power for papers published in high ranked journals. By accounting for quality-adjusted publications, we observe that the observations are more scattered in the DEA vs NDEA (weighted) panel, suggesting greater discrimination from the differences in quality-adjusted publications.

As a final methodological requirement, we test the reliability of our efficiency scores using Spearman's rank correlation test. The motivation for this test is the contention that publications and grants awarded are both inputs as well as outputs. We derive two sets of efficiency scores for this test using only the quality-adjusted

publication indicator and grants awarded. Model 1 considers the publication indicator as the intermediate input and grants awarded as the output and Model 2 uses grants awarded as the intermediate input and the publication indicator as the output. Fig. 3 presents Spearman's rank correlation test for 2004 to 2011. As shown, Spearman's rho ranges from 0.612 to 0.879 with the p-values all less than $\alpha = 0.001$, suggesting a strong positive correlation between the efficiency scores.

What we can infer from the Spearman correlation test is that both the publication indicator and grants awarded can both be both an input as well as an output. Research grants are awarded to research projects with the aim of producing research publications. However, this would mean that academics need to apply and be awarded grants, which is highly unlikely for an academic with no publications. Furthermore, grants are normally awarded to those who have demonstrated some history and consistency in research publications. Thus, it would be more reasonable to believe that grants are awarded to academics with a consistent research track record before being awarded grants.

5. Concluding remarks

This study employed a NDEA model to illustrate a production model that is representative of university operations in the area of research. It is well known that institutions of higher education are complex organizations that use multiple inputs to produce multiple outputs [54]. Due to the nature of higher education organization, in some stages of production, the outputs that are produced using inputs become inputs themselves in the next stage of production. Hence, it is not reasonable to consider a university production model as simply one that transforms inputs to produce outputs. As discussed, the standard DEA model is limited in that it treats the production process like a 'black box' and neglects intervening processes such as different series or parallel functions. In turn, being able to identify stages of production allows a clearer understanding of the (in)efficiencies that may exist when standard DEA fails.

Our findings show that the standard DEA model tends to overstate the efficiency levels of most Australian universities. When we considered the centralized model, we found that universities generally perform better in the first stage, but fare
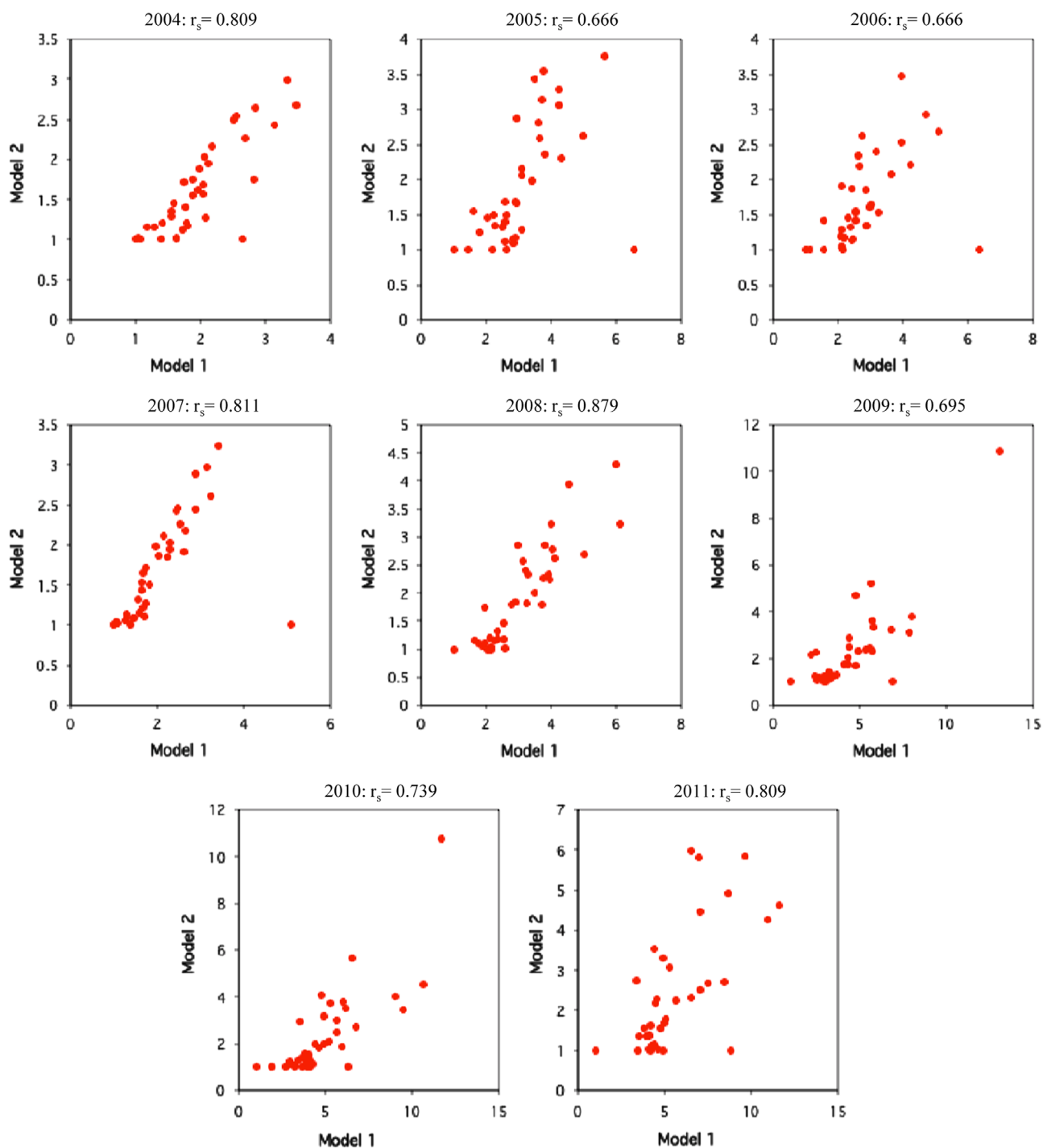
Fig. 3. Spearman rank correlations between efficiency scores.

poorly in the second stage. On this basis, we conclude that it is easier for academics to produce research publications than to win research grants, upon which most Australian academics would surely agree! However, we need to realize that the production model of universities is unique because this particular final output (i.e. total research grants) is not 'produced' per se but is an award granted. This suggests that other factors (both endogenous and exogenous) play a significant role in influencing the outcome of 'total research grants'.

While we only considered the research activities of universities, there is ample room to broaden this approach by including teaching activity. Accounting for teaching and research in the university production model would provide a better and clearer understanding of the network processes in a university and improve resource allocation once we identify any inefficiency. Another extension would be to analyze the sources of inefficiency using a regression-based analysis as common in the efficiency literature. An approach widely used is the bootstrap truncated regression model of [55]. This would quantify the sources of inefficiencies for each of our two stages and provide vital information to policymakers and other stakeholders on organizational weaknesses and directions to improve performance.

References

[1] ACIL Allen Consulting. Benefits Realization Review of Excellence in Research for Australia; 2013.

[2] Herbst M. Financing Public Universities: The Case of Performance Funding. Dordrecht: Springer; 2007.

[3] Johnes G, Johnes J. Measuring the research performance of UK economics departments: application of data envelopment analysis. Oxford Economic Papers 1993;45:332–48.

[4] Abbott M, Doucouliagos C. The efficiency of Australian universities: a data envelopment analysis. Economics of Education Review 2003;22:89–97.

[5] Avkiran NK. Investigating technical and scale efficiencies of Australian universities through data envelopment analysis. Socio-Economic Planning Sciences 2001;35:57–80.

[6] Flegg AT, Allen DO, Field K, Thurlow TW. Measuring the efficiency of British universities: a multi-period data envelopment analysis. Education Economics 2004;12:231–49.

[7] McMillan ML, Chan WH. University efficiency: a comparison and consolidation of results from stochastic and nonstochastic methods. Education Economics 2006;14:1–30.

[8] Flegg AT, Allen DO. Does expansion cause congestion? the case of the older British universities, 1994–2004 Education Economics 2007;15:75–102.

[9] Lee BL. Efficiency of research performance of australian universities: a reappraisal using a bootstrap truncated regression approach. Economic Analysis and Policy 2011;41:195–203.

[10] Färe R. Measuring Farrell efficiency for a firm with intermediate inputs. Academia Economic Papers 1991;19:329–40.

[11] Färe R, Grosskopf S. Intertemporal production frontiers: with dynamic DEA. Boston: Kluwer Academic Publishers; 1996.

[12] Färe R, Grosskopf S. Network DEA. Socio-Economic Planning Sciences 2000;34:35–49.

[13] Adler N, Liebert V, Yazhemsky E. Benchmarking airports from a managerial perspective. Omega 2013;41:442–58.

[14] Yu M-M. Assessment of airport performance using the SBM-NDEA model. Omega 2010;38:440–52.

[15] Akther S, Fukuyama H, Weber WL. Estimating two-stage network Slacks-based inefficiency: an application to Bangladesh banking. Omega 2013;41:88–96.

[16] Avkiran NK. An illustration of dynamic network DEA in commercial banking including robustness tests. Omega 2015;55:141–50.

[17] Fukuyama H, Weber WL. A slacks-based inefficiency measure for a two-stage system with bad outputs. Omega 2010;38:398–409.

[18] Liu W-B, Zhou Z-B, Ma C-Q, Liu D-B, Shen W-F. Two-stage DEA models with undesirable input-intermediate-outputs. Omega 2015;56:74–87.

[19] Matthews K. Risk management and managerial efficiency in Chinese banks: a network DEA framework. Omega 2013;41:207–15.

[20] Wang K, Huang W, Wu J, Liu Y-N. Efficiency measures of the Chinese commercial banking system using an additive two-stage DEA. Omega 2014;44:5–20.

[21] Huang C-W, Ho FN, Chiu Y-H. Measurement of tourist hotels' productive efficiency, occupancy, and catering service effectiveness using a modified two-stage DEA model in Taiwan. Omega 2014;48:49–59.

[22] Kao C. Efficiency decomposition in network data envelopment analysis with slacks-based measures. Omega 2014;45:1–6.

[23] Tone K, Tsutsui M. Network DEA: a slacks-based measure approach. European Journal of Operational Research 2009;197:243–52.

[24] Tone K, Tsutsui M. Dynamic DEA with network structure: A slacks-based measure approach. Omega 2014;42:124–31.

[25] Simon J, Simon C, Arias A. Changes in productivity of Spanish university libraries. Omega 2011;39:578–88.

[26] Li Y-J, Chen Y, Liang L, Xie J-H. DEA models for extended two-stage network structures. Omega 2012;40:611–8.

[27] Liu JS, Lu W-M. DEA and ranking with the network-based approach: a case of R&D performance. Omega 2010;38:453–64.

[28] Foltz JD, Barham BL, Chavas JP, Kim K. Efficiency and technological change at US research universities. Journal of Productivity Analysis 2012;37:171–86.

[29] Valadkhani A, Worthington A. Ranking and clustering Australian university research performance, 1998–2002. Journal of Higher Education Policy and Management 2006;28:189–210.

[30] Worthington A, Lee BL. Efficiency, technology and productivity change in Australian universities, 1998–2003. Economics of Education Review 2008;27:285–98.

[31] Worthington A, Higgs H. Economies of scale and scope in Australian higher education. Higher Education 2011;61:387–414.

[32] Toutkoushian RK, Porter SR, Danielson C, Hollis PR. Using publications count to measure an institution's research productivity. Research in Higher Education 2003;44:121–48.

[33] Johnes J, Yu L. Measuring the performance of Chinese higher education institutions using data envelopment analysis. China Economic Review 2008;19:679–96.

[34] Ng YC, Li SK. Efficiency and productivity growth in Chinese universities during the post-reform period. China Economic Review 2009;20:183–92.

[35] Madden G, Savage S, Kemp S. Measuring public sector efficiency: a study of economics departments at Australian universities. Education Economics 1997;5:153–68.

[36] Stevens PA. A stochastic frontier analysis of English and Welsh universities. Education Economics 2005;13:355–74.

[37] Agasisti T, Johnes G. Beyond frontiers: comparing the efficiency of higher education decision-making units across more than one country. Education Economics 2009;17:59–79.

[38] Johnes G, Johnes J. Higher education institutions' costs and efficiency: taking the decomposition a further step. Economics of Education Review 2009;28:107–13.

[39] Kempkes G, Pohl C. The efficiency of German universities – some evidence from nonparametric and parametric methods. Applied Economics 2010;42:2063–79.

[40] Johnes G, Schwarzenberger A. Differences in cost structure and the evaluation of efficiency: the case of German universities. Education Economics 2011;19:487–99.

[41] Agasisti T, Pohl C. Comparing German and Italian public universities: convergence or divergence in the higher education landscape? Managerial and Decision Economics 2012;33:71–85.

[42] Izadi H, Johnes G, Oskrochi R, Crouchley R. Stochastic frontier estimation of a CES cost function: the case of higher education in Britain. Economics of Education Review 2002;21:63–71.

[43] Rudd E. The evaluation of the quality of research. Studies in Higher Education 1988;13:45–57.

[44] Brinn T, Jones MJ, Pendlebury M. Measuring Research Quality: Peer Review 1, Citation Indices 0. Omega 2000;28:237–9.

[45] Johnes G. Research performance indications in the university sector. Higher Education Quarterly 1988;42:54–71.

[46] Cave M, Hanney S, Kogan M. The use of performance indicators in higher education: a critical analysis of developing practice. London: Jessica Kingsley Publishers; 1991.

[47] González-Pereira B, Guerrero-Bote VP, Moya-Anegón F. A new approach to the metric of journals scientific prestige: the SJR indicator. Journal of Informetrics 2010;4:379–91.

[48] Guz AN, Rushchitsky JJ. Scopus: a system for the evaluation of scientific journals. International Applied Mechanics 2009;45:351–62.

[49] Falagas ME, Kouranos VD, Arencibia-Jorge R, Karageorgopoulos DE. Comparison of SCImago journal rank indicator with journal impact factor. The FASEB Journal 2008;22:2623–8.

[50] Bollen J, van de Sompel H, Hagberg A, Chute R. A principal component analysis of 39 scientific impact measures. PLoS One 2009;4:e6022. http://dx.doi.org/10.1371/journal.pone.0006022.

[51] Liang L, Cook WD, Zhu J. DEA models for two-stage processes: game approach and efficiency decomposition. Naval Research Logistics 2008;55:643–53.

[52] Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision-making units. European Journal of Operations Research 1978;2:429–44.

[53] Kao C, Hwang SN. Efficiency decomposition in two-stage data envelopment analysis: an application to non-life insurance companies in Taiwan. European Journal of Operational Research 2008;185:418–29.

[54] Monfared MAS, Safi M. Network DEA: an application to analysis of academic performance. Journal of Industrial Engineering International 2013;9:1–10.

[55] Simar L, Wilson P. Estimation and inference in two-stage, semi-parametric models of productions processes. Journal of Econometrics 2007;136:31–64.