Journal of Informetrics 10 (2016) 238–253

Contents lists available at ScienceDirect

Journal of Informetrics

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

Using multi-level frontiers in DEA models to grade countries/territories

Guoliang Yang[a], Per Ahlgren[b], Liying Yang[c,*], Ronald Rousseau[d,e], Jielan Ding[c]

[a] Institute of Policy and Management, Chinese Academy of Sciences, Beijing 100190, China
[b] School of Education and Communication in Engineering Sciences (ECE), KTH Royal Institute of Technology, Stockholm 100 44, Sweden
[c] National Science Library, Chinese Academy of Sciences, Beijing 100190, China
[d] Institute for Education and Information Sciences, IBW, University of Antwerp (UA), Antwerp B-2000, Belgium
[e] KU Leuven, Department of Mathematics, Leuven B-3000, Belgium

ARTICLE INFO

Article history:
Received 20 September 2015
Received in revised form
20 November 2015
Accepted 12 January 2016
Available online 1 February 2016

Keywords:
Countries/territories
Grades
Data envelopment analysis (DEA)
Efficient frontier

ABSTRACT

Several investigations to and approaches for categorizing academic journals/institutions/countries into different grades have been published in the past. To the best of our knowledge, most existing grading methods use either a weighted sum of quantitative indicators (including the case of one properly defined quantitative indicator) or quantified peer review results. Performance measurement is an important issue of concern for science and technology (S&T) management. In this paper we address this issue, leading to multi-level frontiers resulting from data envelopment analysis (DEA) models to grade selected countries/territories. We use research funding and researchers as input indicators, and take papers, citations and patents as output indicators. Our research results show that using DEA frontiers we can unite countries/territories by different grades. These grades reflect the corresponding countries' levels of performance with respect to multiple inputs and outputs. Furthermore, we use papers, citations and patents as single output (with research funding and researchers as inputs), respectively, to show country/territory grade changes. In order to increase the insight in this approach, we also incorporate a simple value judgment (that the number of citations is more important than the number of papers) as prior information into the DEA models to study the resulting changes of these Countries/Territories' performance grades.

1. Introduction

Several efforts have been made regarding categorization of countries, industries, research institutions and academic journals into different grades of standing or quality. Yang, Zhou, and Yue (2013b) analyzed the overall development and the balance of the disciplinary structure of China's science based on papers covered by the Science Citation Index and using bibliometric methods. These authors further categorized selected countries to reflect their developmental status. Hatzichronoglou (1997) introduced four categories of industries (high-, medium-high, medium-low and low technology) based on direct R&D intensity as well as R&D embodied in intermediate and investment goods. In 2005, the Chinese Academy of Sciences (CAS) evaluated its subordinate institutes and classified them into three grades (Excellent, Good, and Satisfactory)

* Corresponding author. Tel.: +86 10 82627304.
E-mail addresses: glyang@casipm.ac.cn (G. Yang), perahl@kth.se (P. Ahlgren), yangly@mail.las.ac.cn (L. Yang), ronald.rousseau@kuleuven.be (R. Rousseau), dingjielan@mail.las.ac.cn (J. Ding).

http://dx.doi.org/10.1016/j.joi.2016.01.008
1751-1577/

(CAS, 2006). From 2007 on, the Association of Business Schools (ABS) started to issue the Academic Journal Quality Guide, which classifies journals in business and management into four categories (grade 1 to 4) recognizing the quality of those journals based on a survey of hundreds of experts in this field (Harvey, Morris, & Kelly, 2007a, 2007b, 2008). From 2010 on a new category, termed 4*, was added to the existing categories in the ABS system to recognize the quality of the top journals (Harvey, Kelly, Morris, & Rowlinson, 2010). Bandyopadhyay (2013) categorized business and management journals into four categories (Excellent, Very Good, Standard, Satisfactory) based on multiple inputs, including ISI's Social Science Citation lists of ranked journals and ISI's impact factor analyses. Glänzel (2011) used characteristic scores and scales as parameter-free tools to identify top journals.

The grading methods in the research reported above use either a weighted sum of quantitative indicators (including the case of one properly defined quantitative indicator) or quantified peer or expert review results. In general, the weighted sum approach normally needs to know the weights of indicators and corresponding threshold values as a priori information, while the peer or expert review process usually costs a lot of time and money (Smith, 1996).

In the light of these issues, this paper presents an alternative approach, involving multi-level DEA (Data Envelopment Analysis) frontiers, to divide various countries/territories into different grades. DEA is a nonparametric method and one of the most popular mathematical tools for estimating the relative efficiency of decision-making units (DMUs), which has been used earlier in scientometrics and informetrics, including research efficiency of countries (e.g., Rousseau & Rousseau, 1997, 1998). One of the main advantages of DEA approach is to allow the DMUs to freely select their weights for the indicators that are most favorable with respect to achievement of maximum performance. The DMUs on the frontier will be assessed as efficient, which are considered to represent the best practices and have the same level of performance. Those DMUs that are not on the frontier will be compared with their peers or projections on the frontier to measure their relative performance. The flexibility of weight selection by this mathematical model makes it unnecessary to know the weights of indicators and corresponding threshold values in advance. Moreover, there is no need to obtain information from a peer or expert review process when applying DEA assessment, so resources can be saved in terms of costs and time. Different from the normal application of DEA approach, however, in this paper we use DEA in a novel way to grade selected countries/territories instead of measuring their efficiencies. The paper is a revised and expanded version of (Yang, Ahlgren, Yang, Rousseau, & Ding, 2015).

The rest of the paper is organized as follows. Section 2 gives a literature review. Section 3 introduces the input and output indicators, and the corresponding dataset used in the analysis. The used methods are described in Section 4, in which we treat multi-level efficient frontiers and show how to subdivide countries/territories into different grades using these frontiers. In Section 5, the results of the study are given, whereas conclusions appear in Section 6.

2. Literature review

DEA is a nonparametric method used in operations research and economics for the estimation of production frontiers. It is mainly used to empirically measure the productive efficiency of decision making units (or DMUs) (Charnes, Cooper, & Rhodes, 1978). The efficiency of science and technology (S&T) resource utilization is an important issue in S&T management (Yang, Yang, Liu, Li, & Fan, 2013a; Yang, Rousseau, Yang, & Liu, 2014a). Johnes and Johnes (1992) evaluated the efficiency of S&T organizations using data envelopment analysis (DEA) as a performance analysis tool. Rousseau & Rousseau (1997, 1998) assessed the efficiency of countries using gross domestic product, active population and research and development (R&D) expenditure as inputs, and publications and patents as outputs. They showed that DEA can be used in scientometrics as a tool to measure the efficiency of decision making units (DMUs, e.g., countries) by gauging closeness to the efficiency frontier. Similar techniques have been used by other researchers (Kao & Lin, 1999; Roy & Nagpaul, 2001; Shim & Kantor, 1998). Yang and Chang (2009) used DEA under constant and variable returns to scale (RTS) to measure firms' efficiency. Worthington (2001) conducted an empirical survey of frontier efficiency measurement techniques in education. Other researchers analyzed the efficiency or productivity in the education sector, e.g., Abbott and Doucouliagos (2003), Avkiran (2001), Carrington, Coelli, and Rao (2005), Worthington and Lee (2008), Flegg, Allen, Field, and Thurlow (2004), Johnes and Johnes (1992), Kempkes and Pohl (2010), Wolszczak-Derlacz and Parteka (2011), and Aristovnik (2012). Recently, Grosskopf, Hayes, and Taylor (2014) summarize some of the modeling and measurement choices that are important for those seeking to apply frontier efficiency methods to the educational arena. When studying the standard university model, Brandt and Schubert (2013) observed that universities are large agglomerations of many (often loosely affiliated) small research groups. They explained this observation by typical features of the scientific production process. In particular, they argued that there are decreasing returns to scale (RTS) on the level of individual research groups. RTS is a concept with strong relation to scale efficiency. Somewhat similar observations, relating to decreasing RTS, were published earlier by Bonaccorsi and Daraio (2005). Schubert (2014) used non-parametric techniques of multidimensional efficiency measurement (e.g., DEA) to analyze RTS in scientific production based on survey data for German research groups from three scientific fields. Based on DEA models, Yang et al. (2013a, 2014a[1]) analyzed the directional RTS of biological institutes in the CAS. On the assessment of RTS in higher education (European universities), Daraio, Bonaccorsi, and Simar (2015) propose a new approach by taking scale effects and specialization into account.

[1] *Note:* There is a typo in Yang et al. (2014a), where a free variable $-\mu_p$ is missing in the left side of the third constraint in model (4). Consequently, in its equivalent Model (5) there also should be a free variable $-\mu_p'$ in the third constraint. This error does not affect the results and discussions in their paper.

Some fairly recent studies examined the efficiency of countries or regions in utilizing R&D expenditures or other resources. Lee and Park (2005) evaluated R&D efficiency across nations using patents, technology balance of receipts and journal articles as outputs. Wang and Huang (2007) analyzed R&D efficiency of nations by considering patents and papers as outputs. Lee, Park, and Choi (2009) used DEA to measure and compare the performance of national R&D programs in South Korea. Sharma and Thomas (2008) investigated the R&D efficiency of developing countries in relation to developed countries, taking into account time lags. Similar investigations such as Chen, Hu, and Yang (2011) measured the R&D efficiency of multiple innovation outputs in a national innovation system. Zhong, Yuan, Li, and Huang (2011) conducted a performance evaluation of regional R&D investments in China based on the first official economic census data for China. Aristovnik (2012) assessed the relative efficiency of education and R&D expenditures in the new EU member states. Sueyoshi and Goto (2013) used a DEA–DA (Data Envelopment Analysis–Discriminant Analysis) approach to measure the importance of R&D expenditure in Japanese information technology industry. The literature referred to hitherto focused on the quantitative measurement of efficiency of resource utilization. Banker (1993) pointed out that DEA efficiency scores usually overestimate the efficiency and are biased. Smith (1997) argued that the extent of the overestimation is highly dependent on sample size and the complexity of the production process (as indicated by the numbers of inputs and outputs).

However, in many cases a full ranking of DMUs is not needed or even meaningless, and it suffices to know the performance category or grade to which a particular DMU belongs. Although DEA is originally a mathematical tool for estimating relative efficiency and has a strong link to production theory in economics, the tool can also be used for benchmarking in operations management, where a set of measures is selected to benchmark the performance of manufacturing and service operations. In the circumstance of benchmarking, the efficient DMUs, as defined by DEA, may not necessarily form a “production frontier”, but rather lead to a “best-practice frontier” (Cook, Tone, & Zhu, 2014). Hereafter we still use the term ‘efficient frontier’ to denote the frontier produced from DEA models in this paper. In this case DEA is used to detect the “best practitioners” as the benchmark instead of relative efficiency. The issue becomes one of how to classify these performance measures into inputs and outputs, for use in DEA. In general, DEA minimizes “inputs” and maximizes “outputs”, i.e., smaller levels of the former and larger levels of the latter represent better performance. This can then be a rule for classifying factors. Therefore this paper does not aim to measure the efficiency of these countries/territories. Instead it aims to use the DEA model as a mathematical tool to benchmark the performance of countries/territories and grade them into different performance levels.

3. Indicators and data

In this study, research funding and researchers are used as input indicators. The time period of data collection is from Jan. 2008 to Dec. 2012. Research funding here means the total Gross Domestic Expenditure on R&D (in millions of US dollars, using Purchasing Power Parity (PPP$)) in this period. The researcher indicator measures the number of total researchers (Full time equivalents, FTEs) in a country. As output indicators, we select papers, citations and patents, which are frequently used in the existing literature (e.g., Rousseau & Rousseau, 1997, 1998; Yang et al., 2013b). Sharma and Thomas (2008) took into account time lags when investigating the R&D efficiency of developing countries in relation to developed countries. In order to eliminate the influence of time lags between inputs and outputs, we use the total number of papers, where the papers are covered by the Science Citation Index (SCI) and the Social Science Citation Index (SSCI) and published in the period Jan. 2008–Dec. 2012. The citation indicator refers to the total number of citations received by the papers published in the five-year period 2008–2012, whereas the citations to these papers were collected July 22 2015. The Patents indicator refers to the total number of authorized patents of each country in the period Jan. 2008–Dec. 2012.

We use OECD statistics[2], Thomson Reuters' research evaluation tool InCites[3], and the World Intellectual Property Organization (WIPO) statistics[4] as sources for input and output data, respectively. All 34 OECD member countries and seven non-OECD member countries/territories were selected for the study. Other countries/territories, namely Australia, Greece, Iceland, Israel, Mexico, New Zealand, Switzerland, and Taiwan, although generally covered by OECD statistics, were excluded due to lack of input data for one or several years. For this reason the number of selected countries is only 33. See Table 1 for details.

4. Methods

4.1. DEA models and their frontiers

DEA is an approach based on linear programming for analyzing performance of organizations and operational processes. The approach was first proposed by Charnes et al. (1978). The DEA models can use input and output data to evaluate the relative efficiency or performance of DMUs without prior knowledge of input/output functions and of indicator weights (Cook et al., 2014). Nowadays, numerous theoretical and empirical works on the DEA method have been published, extending the

[2] http://www.oecd-ilibrary.org/statistics.
[3] http://incites.isiknowledge.com/Home.action (note that this URL can be accessed only when having a subscription to InCites).
[4] http://ipstats.wipo.int/ipstatv2/index.htm?tab=patent.

Table 1
Values of input and output indicators across 33 countries/territories between 2008 and 2012.

No.	Countries/	Output			Input	
	Territories	Papers	Citations	Patents	Research funding (PPP)	Researcher (FTE)
1	Argentina	38,258	369,438	1043	19,866.44	234,758
2	Austria	60,189	885,749	22,117	48,117.08	182,228.6
3	Belgium	89,372	1403,789	24,827	44,696.87	202,331
4	Canada	286,704	4066,683	49,453	125,482.8	785,720
5	Chile	26,461	253,761	1016	5594.19	29,134.49
6	China	726,594	6942,097	475,730	1082,947	6677,675
7	Czech Republic	46,839	481,928	3483	21,025.38	151,671.2
8	Denmark	62,639	1092,665	19,181	34,284.6	190,051.8
9	Estonia	6833	87,679	408	2639.39	21,463
10	Finland	51,831	763,451	30,540	37,991.71	203,623.7
11	France	330,123	4504,676	170,070	255,004.3	1213,890
12	Germany	460,595	6731,020	354,505	449,596.3	1639,052
13	Hungary	29,214	325,981	2884	12,452.32	106,766
14	Ireland	33,955	493,536	8289	15,394.13	73,771
15	Italy	268,105	3610,272	97,616	126,495.4	517,876.9
16	Japan	385,324	4081,653	1431,311	726,479.9	3271,236
17	Luxembourg	2800	34,301	4282	3240.87	12,841.6
18	Netherlands	160,853	2792,128	73,217	67,467.68	285,974.9
19	Norway	50,105	691,392	11,086	24,504.85	133,371
20	Poland	101,778	771,858	9360	28,960.3	318,554.9
21	Portugal	49,106	582,132	1459	20,775.59	208,319.7
22	Romania	33,424	191,870	2605	8341.05	92,541
23	Russia	141,537	774,600	122,441	171,786.5	2226,395
24	Singapore	46,349	708,136	9086	38,424.03	158,261.3
25	Slovakia	15,134	122,950	615	4034	71,656.9
26	Slovenia	17,662	160,035	2262	6082.36	39,839
27	South Africa	42,500	421,899	6438	24,011.12	99,394.5
28	South Korea	209,208	1919,110	429,325	264,904.3	1348,822
29	Spain	236,216	2903,644	23,697	100,907.9	656,455.1
30	Sweden	103,976	1638,689	52,078	65,700.14	244,561
31	Turkey	114,351	693,814	4164	50,140.48	329,141.4
32	UK	486,505	7463,557	84,674	194,953.5	1272,155
33	USA	1773,100	27,267,062	930,377	2088,559	6158,300

Data sources: Inputs: OECD statistics. http://www.oecd-ilibrary.org/statistics. Outputs: (1) InCites. http://incites.isiknowledge.com/Home.action; (2) WIPO. http://ipstats.wipo.int/ipstatv2/index.htm?tab=patent.

original approach in different ways, and applying them to many areas, including the private and the public sector (e.g., Cooper, Seiford, & Tone, 2007).

Let $X = (x_1, x_2, \ldots, x_m)$ and $Y = (y_1, y_2, \ldots, y_s)$ be input and output vectors of m and s dimensions, respectively, of n DMUs. Then the Production Possibility Set (PPS) is defined by

$$\mathrm{PPS} = \left\{(X, Y) : X \ \text{ can } \ \text{produce} \ \ Y\right\} \tag{1}$$

There can be different forms of PPS based on different assumptions. The PPS implied in the CCR (Charnes, Cooper, and Rhodes)-DEA model, which was proposed by Charnes et al. (1978) under the assumption of constant RTS, is defined as follows:

$$\mathrm{PPS}\,(X, Y) = \left\{ (X, Y)\,|X \geq \sum_{j=1}^{n} \lambda_j X_j,\ Y \leq \sum_{j=1}^{n} \lambda_j Y_j,\ \lambda_j \geq 0, \quad j = 1, \ldots, n \right\} \tag{2}$$

Banker, Charnes, and Cooper (1984) defined the PPS under the assumption of variable RTS to obtain the BCC (Banker, Charnes, and Cooper)–DEA model:

$$\mathrm{PPS}\,(X, Y) = \left\{ (X, Y)\,|X \geq \sum_{j=1}^{n} \lambda_j X_j,\ Y \leq \sum_{j=1}^{n} \lambda_j Y_j,\ \sum_{j=1}^{n} \lambda_j = 1,\ \lambda_j \geq 0, \quad j = 1, \ldots, n \right\} \tag{3}$$

where $\lambda_j\,(j = 1, \ldots, n)$ are coefficients, and $X_j = \left(x_{1j}, x_{2j}, \ldots, x_{mj}\right) \in \mathbb{R}_m^+$ and $Y_j = \left(y_{1j}, y_{2j}, \ldots, y_{sj}\right) \in \mathbb{R}_s^+$ denote the input and output vectors of $\mathrm{DMU}_j\,(j = 1, \ldots, n)$. The boundary of the PPS is referred to as the production frontier.

Remark 1. RTS is an economic concept describing the relationship between changes in the scale of inputs and outputs. The variable RTS assumption reflects the case where the production technology may exhibit increasing, constant or decreasing

RTS, and the constant RTS assumption reflects the case where the outputs will change by the same proportion as the inputs (e.g. doubling of all inputs will double the outputs).

Definition 1. The efficient frontier of PPS is defined as follows:

$$\mathrm{EF} = \left\{(X, Y) \in PPS | \text{there is no } \left(\bar{X}, \bar{Y}\right) \in \mathrm{PPS} \text{ such that } \left(-\bar{X}, \bar{Y}\right) > (-X, Y)\right\} \tag{4}$$

Note: This unobservable production frontier is called the true frontier or true efficient frontier hereafter. When there is only a single output, the production frontier is called production function in the economic literature. DMUs which are technically efficient operate on the frontier, while technically inefficient DMUs operate at points in the interior of the PPS. Thus, it is rational to rank DMUs according to their distances to the true frontier.

The core idea of the classic DEA is to identify first the production frontier. DMUs on the frontier are regarded as efficient. Those DMUs not on the frontier will be compared with their peers or projections on the frontier to measure their relative efficiencies. All the DMUs on the frontier are considered to represent the best practices and have the same level of performance.

Based on existing observations DMU_j ($j = 1, \ldots, n$), DEA models construct a piecewise linear production frontier, a non-parametric estimate of the unobservable true frontier. Then DEA models measure the efficiency or performance of a DMU via its distance to the estimated frontier. Using radial measurement and input orientation, we have the following input-based CCR-DEA model (Charnes et al., 1978):

$$\theta_c^* = \min\theta$$
$$s.t. \begin{cases} \sum_{j=1}^{n} x_{ij}\lambda_j \le \theta x_{i0}, & i = 1, \ldots, m \\ \sum_{j=1}^{n} y_{rj}\lambda_j \ge y_{r0}, & r = 1, \ldots, s \\ \lambda_j \ge 0, & j = 1, \ldots, n \end{cases} \tag{5}$$

where $\lambda_j \geq 0$ are the multipliers of inputs and outputs. Here θ_c^* measure the degree of efficiency or performance by radial measurement under the assumption of constant RTS.

If we assume that the production technology satisfies variable returns to scale assumption, we have the following input-based BCC–DEA model (Banker et al., 1984):

$$\theta_b^* = \min\theta$$
$$s.t. \begin{cases} \sum_{j=1}^{n} x_{ij}\lambda_j \le \theta x_{i0}, & i = 1, \ldots, m \\ \sum_{j=1}^{n} y_{rj}\lambda_j \ge y_{r0}, & r = 1, \ldots, s \\ \sum_{j=1}^{n} \lambda_j = 1 & \\ \lambda_j =\ge 0, & j = 1, \ldots, n \end{cases} \tag{6}$$

where θ_b^* measure the degree of efficiency or performance by radial measurement under the assumption of variable returns to scale. It should be noted that Model (6) differs from Model (5) only regarding the constraint $\sum_{j=1}^{n} \lambda_j = 1$, which yields that the variable RTS assumption is satisfied. Obviously, if $\theta_c^* = 1$ in Model (5) or $\theta_b^* = 1$ in Model (6), then the DMU is on the efficient frontier in CCR-DEA and BCC–DEA, respectively.

We visualize the frontier of a DEA model in Fig. 1, using two inputs (x_1 and x_2) and one output (y). The piecewise linear line ABCD defines the efficient frontier of the existing observations. For example, for point *G*, representing a DMU, its efficiency or performance score can be calculated as the ratio of distance OG' to the distance OG.

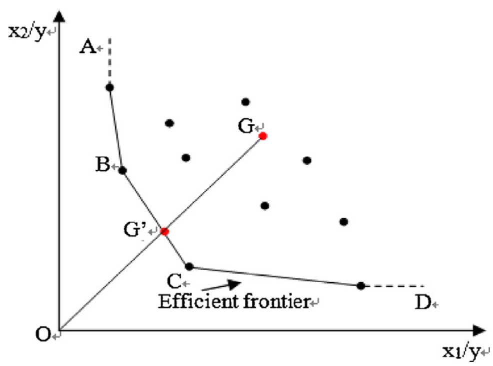

Fig. 1. Efficient Frontier of a DEA model.

Table 2
6 DMUs with 2 inputs and a single output.

DMUs	DMU_1	DMU_2	DMU_3	DMU_4	DMU_5	DMU_6
Output (y)	120	8	24	40	120	24
Input 1 (x_1)	19	1	1	2	10	8
Input 2 (x_2)	10	1	6	15	17	1

Table 3
Expanded DMUs with 2 inputs and single output.

DMUs	DMU_1	DMU_2	DMU_3	DMU_4	DMU_5	DMU_6
Output(y)	120	120	120	120	120	120
Input 1(x_1)	19	15	5	6	10	40
Input 2(x_2)	10	15	30	45	17	5

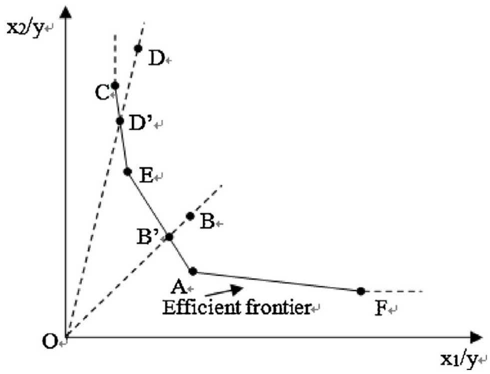

Fig. 2. Efficient Frontier and DMUs.

We now give an example to illustrate the detection of the efficient frontier and the evaluation of DMUs using a DEA model. We suppose there are six DMUs with two inputs and a single output. In Table 2, hypothetical data is given.

First, for comparison, we expand the inputs and output of each DMU proportionally and let the output of each DMU be 120 (Table 3).

We show these six DMUs in Fig. 2 (which gives projections in input space) using points *A*–*F* to denote DMU_1–DMU_6.

We use a piecewise line to link points *C*, *E*, *A*, *F* and merge it with the horizontal and vertical lines from point *F* and *C*, respectively, to obtain the piecewise linear convex hull, which is the efficient frontier produced from DEA model. Points *C*, *E*, *A*, *F* are on the efficient frontier and their efficiencies are all unity. On the contrary, points *B* and *D* are inside the convex hull, so these two DMUs are inefficient compared with their peers or projections (points B' and D') on the efficient frontier. Taking point *B* as example, the DEA model uses the ratio of distance OB' to the distance OB to measure point B's relative efficiency or performance.

The issue of taking value judgments into account frequently arises in both theoretical investigations and practical applications of DEA. The value judgments can be considered as logical constructs and incorporated within an efficiency or performance assessment study to reflect prior views or information (Allen, Athanassopoulos, Dyson, & Thanassoulis, 1997). Therefore more constraints can be added into DEA models (Model (5) or Model (6)) to reflect prior views. See Section 4.4 for more details on this issue.

4.2. Decomposition of countries/territories based on multi-level frontiers in DEA

In Section 3.1, we showed how the efficient frontier can be detected. If we remove the efficient DMUs on the frontier, we can use the DEA model again to obtain a new frontier. We do this repeatedly in order to decompose DMUs into different grades. This process is illustrated in Fig. 3. In this figure, the efficient frontier—grade 1 is the piecewise line ABCD, on which the DMUs with the best level of efficiency or performance are located. After we remove the DMUs on the efficient frontier—grade 1, we rerun the DEA model, obtaining the DMUs on the efficient frontier—grade 2 as the second group, and so on. This process is iterated until there is no DMU left, and then the grading of the DMUs will end. The efficient frontier in Fig. 1 is the same as the efficient frontier—grade 1 in Fig. 3.

In earlier works, DEA frontiers have been used either to measure the relative efficiency or performance of the DMUs (e.g., Charnes et al., 1978; Cook & Seiford, 2009) by comparing them with their peers or projections on the frontier, or to estimate the RTS by the frontier's shape (e.g., Banker, Cooper, Seiford, Thrall, & Zhu, 2004). Seiford and Zhu (2003) proposed context-dependent DEA, which measures the attractiveness when DMUs exhibiting poorer performance are chosen as the evaluation

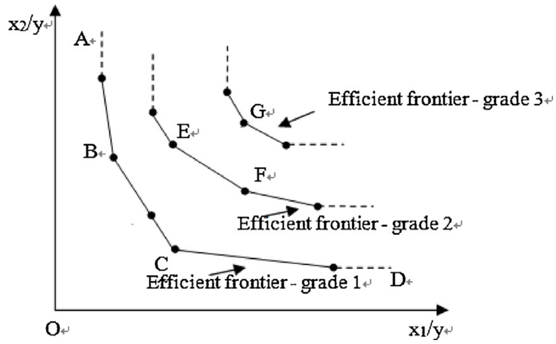

Fig. 3. Multi-level efficient frontiers of a DEA model.

context and the progress when DMUs exhibiting better performance are chosen as the evaluation context. Inspired by their work, we use multi-level frontiers in DEA models to decompose DMUs into different levels to reflect different grades of performance. Furthermore we also use multi-level frontiers to form a performance improvement chain, and we incorporate value judgments or preferences into the construction of the DEA model through proper constraints.

In the process of decomposing the DMUs into different performance grades, denoted as $F_k\,(k = 1, 2, \ldots)$, we need to ensure that a given DMU can only be assigned to one grade to avoid conflicts. An efficient frontier is a convex hull. This implies that if a point belongs to F_{k+l} it cannot belong to any other F_{k+l} (if it exists, where l is a positive integer). Indeed a point on the frontier is a convex linear combination of efficient points on the frontier. If point P would belong to F_{k+l} and F_{k+l} this would mean that P is a convex linear combination of points that do not belong to F_k, which is not possible.

Theorem 1. Let F_k and F_{k+l}, where l is a positive integer, be efficient frontier grades. Then it holds that there is no intersection between F_k and F_{k+l}.

Proof. Let F_k and F_{k+l}, where l is a positive integer, be efficient frontier grades. Assume that there is at least one intersection between F_k and F_{k+l}. We denote the intersection as P_l, which is on the F_k and F_{k+l}. If P_l is on F_{k+l}, its efficiency or performance score with respect to F_k is less than unity, which contradicts the fact that P_l is also on the F_k, from which it follows that its efficiency or performance is unity for this frontier. Q.E.D.

4.3. Identifying the performance targets of countries/territories

In the above Section 4.2 we showed how to decompose the countries/territories into different performance grades based on multi-level frontiers in DEA. In this subsection we will show how to identify the performance targets of countries/territories on each grade. Normally we think the improvement of performance is a step-by-step process which means the countries/territories on one grade should find performance targets on next adjacent grade with higher performance level. Fig. 4 illustrates this process of identifying performance targets. For example, we can see that point G lies on the frontier—grade 3. It is more realistic for this DMU to set a performance target on frontier—grade 2 first instead of one on frontier—grade 1, because the performance level can only be improved gradually. Therefore, point G' on frontier–grade 2 is a practical performance target for point G. In Fig. 4, point G' is the linear combination of points E and F. Thus the performance target point G' for point G on frontier–grade 3 is points E and F with their own weights, i.e., point G' can be formulated

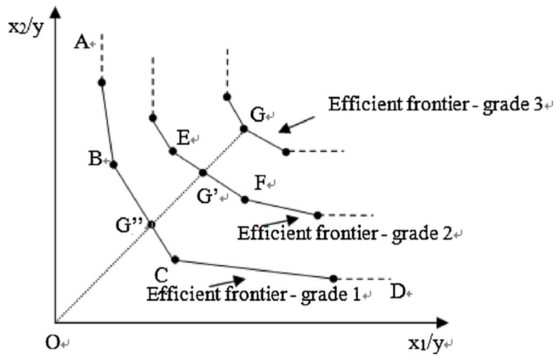

Fig. 4. Illustration of identifying performance targets.

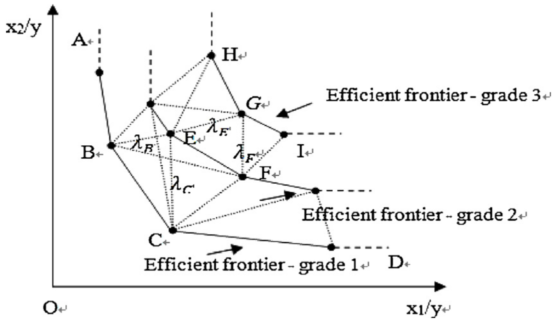

Fig. 5. Illustration of performance improvement chain.

by weighted linear combination of points E and F. It should be noted that countries/territories on the grade 1 don't have performance targets because they have the highest performance already.

Procedure 1 described in the box below reflects this process of identifying performance targets. Assume there are K performance grades, denoted by $F_k\,(k = 1, 2, \ldots, K)$. We start from the grade with the lowest performance level. Then we have:

Procedure 1:
For $k = K$ downto 2 do
Begin
Step1: Run DEA Model (5) or (6) on the dataset $G_k \cup G_{k-1}$, where G_k denotes the set of DMUs on performance grade F_k.
Step2: Record the coefficients λ_j for each DMU on the performance grade F_k as the weights of linear combination of DMUs on grade F_{k-1}.
End;

In Fig. 5, a performance improvement chain is illustrated:

From Fig. 5, we can see that the performance target of point G is the combination of points E and F with weights λ_E and λ_F respectively. The performance targets of points E and F are the combinations of points B and C with weights λ_B and λ_C, respectively. Furthermore we can also set the weights based on the coefficients λ_j in DEA Model (5) or (6) for the performance improvement chain. We will show this in detail in Section 5.

4.4. Incorporating value judgment as prior information into DEA

One may take the standpoint that one indicator is more important than others as the prior information of value judgment. Therefore, in order to increase our insight, we can incorporate this value judgment into the DEA models to show the changes in grade of the countries/territories. Taking BCC–DEA as an example, we can rebuild the BCC–DEA model with the prior information of decision makers' value judgment. First, we consider the dual model of Model (6):

$$s.t.\begin{cases} \max_{u_r,v_i,\mu_0} \sum_{r=1}^{s} u_r y_{r0} + \mu_0 \\ \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + \mu_0 \le 0, \quad j = 1, \ldots, n \\ \sum_{i=1}^{m} v_i x_{i0} = 1 \\ u_r \ge 0, v_i \ge 0, r = 1, \ldots, s, \quad i = 1, \ldots, m, \quad \mu_0 \ \text{free} \end{cases} \tag{7}$$

where u_r and v_i are multipliers, and μ_0 is a variable free of sign.

Remark 2. A linear program is a formulation of an optimization problem: a minimization or maximization of an objective function over some domain. The objective function is linear, and the domain, or feasible set, is defined by linear constraints. The linear program starts with what is typically called the "primal". To each linear program there is associated another linear program called its "dual". Dual of the dual is (equivalent to) the primal. The optimal values of the primal model and dual model are the same. We refer the interested reader to, e.g., Strayer (1989).

Second, we incorporate prior information, transforming Model (7) into Model (8), in which constraints $u_r, v_i \in \Omega$, where Ω denotes the set of preferential and cognitive constraints on input and output weights, are added to ensure that the prior value judgments can be reflected.

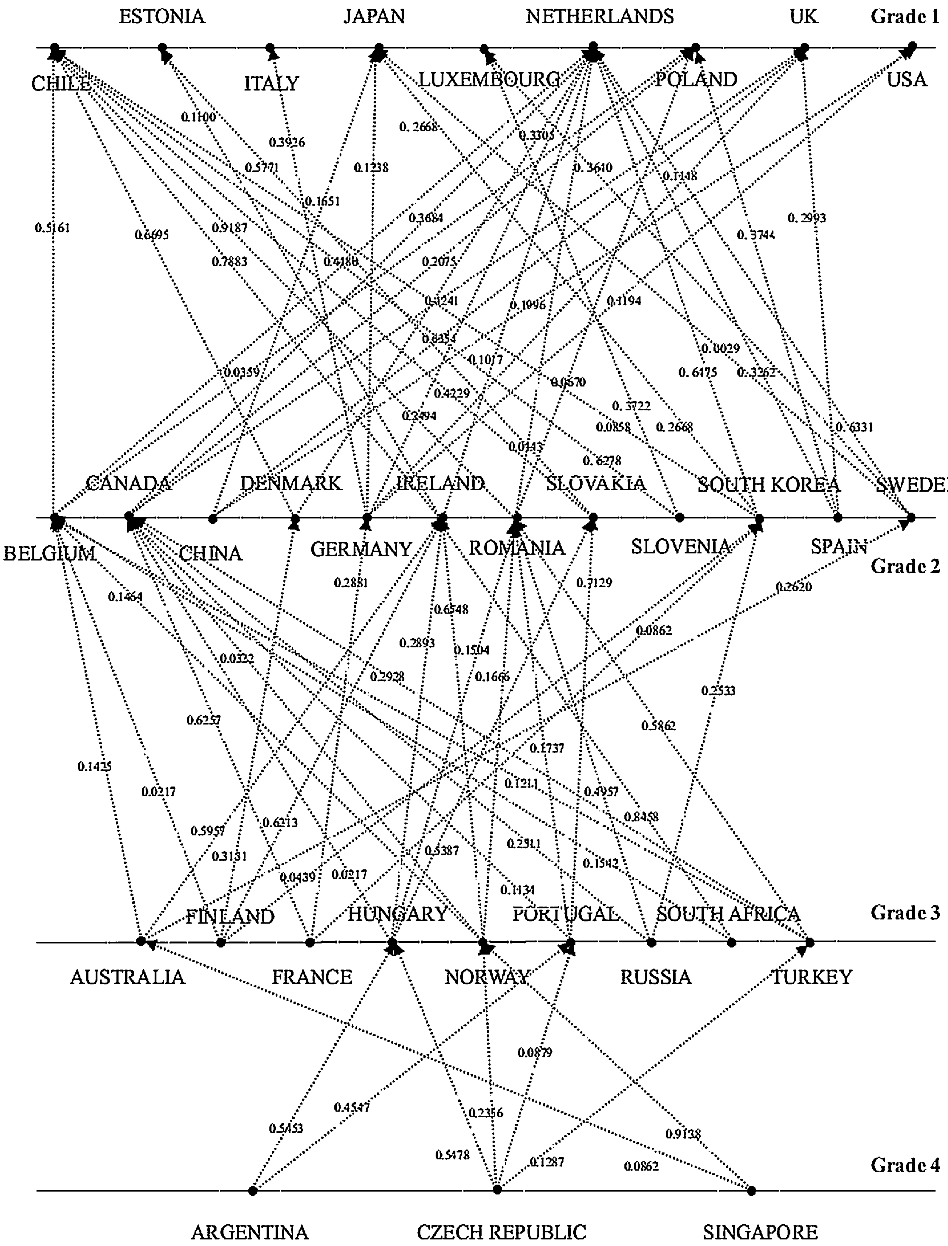

Fig. 6. The performance improvement chain of the countries/territories.

$$s.t.\begin{cases} \max_{u_r,v_i,\mu_0}\sum_{r=1}^{s}u_r y_{r0}+\mu_0 \\ \sum_{r=1}^{s}u_r y_{rj}-\sum_{i=1}^{m}v_i x_{ij}+\mu_0\le 0, \quad j=1,\ldots,n \\ \sum_{i=1}^{m}v_i x_{i0}=1 \\ u_r, v_i \in \Omega \\ u_r\ge 0, v_i\ge 0, \quad r=1,\ldots,s, \quad i=1,\ldots,m, \quad \mu_0 \ \text{free} \end{cases} \tag{8}$$

It should be noted that Ω denotes the relation set of coefficients u_r, v_i of inputs and outputs, which could be determined according to the real practices. For example, if the prior information is that one output indicator Y_a is more important than another output indicator Y_b, we can let the set $\Omega=\{u_a\ge u_b\}$ in Model (8). We used Model (8) to obtain additional

Table 4
Grades of the countries/territories.

No.	Countries/ Territories	Grades (two inputs and three outputs)	Grades (two inputs and the first output (papers))	Grades (two inputs and the second output (citations))	Grades (two inputs and the third output (patents))
1	DMU1-Argentina	4	4	5	5
2	DMU 2-Austria	3	4	4	5
3	DMU 3-Belgium	2	2	2	5
4	DMU 4-Canada	2	2	2	6
5	DMU 5-Chile	1	1	1	2
6	DMU 6-China	2	2	2	3
7	DMU 7-Czech republic	4	4	5	5
8	DMU 8-Denmark	2	4	3	5
9	DMU 9-Estonia	1	1	1	1
10	DMU 10-Finland	3	5	5	4
11	DMU 11-FrancE	3	3	3	4
12	DMU 12-Germany	2	2	2	3
13	DMU 13-Hungary	3	3	4	4
14	DMU 14-Ireland	2	2	2	3
15	DMU 15-Italy	1	1	2	4
16	DMU 16-Japan	1	3	4	1
17	DMU 17-Luxembourg	1	1	1	1
18	DMU 18-Netherlands	1	1	1	3
19	DMU 19-Norway	3	3	3	5
20	DMU 20-Poland	1	1	4	6
21	DMU 21-Portugal	3	3	4	6
22	DMU 22-Romania	2	2	3	3
23	DMU 23-Russia	3	5	5	5
24	DMU 24-Singapore	4	5	4	6
25	DMU 25-Slovakia	2	2	2	2
26	DMU 26-Slovenia	2	2	2	2
27	DMU 27-South Africa	3	3	3	4
28	DMU 28-South Korea	2	4	4	2
29	DMU 29-Spain	2	2	3	7
30	DMU 30-Sweden	2	2	3	4
31	DMU 31-Turkey	3	3	6	7
32	DMU 32-UK	1	1	1	5
33	DMU 33-USA	1	1	1	2

performance grading results. Moreover, when incorporating value judgment as prior information into DEA, we can use the dual model of Model (8) to find coefficient vectors to reflect the performance targets of each DMU and form a performance improvement chain graph, as illustrated in Section 4.3.

4.5. The pros and cons of the DEA approach

As a nonparametric method, DEA has been widely used in productivity or performance evaluation and in the efficiency analysis of many business and non-profit organizations. The main idea of the DEA is first to identify the production frontier and then, the DMUs on the frontier will be assessed as efficient. Those DMUs that are not on the frontier will be compared with their peers or projections on the frontier to measure their relative efficiencies. All of the DMUs on the frontier are considered to represent the best practices and, within this framework, are considered to have the same level of performance. One of the main advantages of DEA approach is to allow the DMUs to freely select the indicator weights that are most favorable for them regarding achievement of maximum performance. However, the DEA approach also has some well-known drawbacks, namely its deterministic nature and the “curse” of dimensionality. The deterministic nature of DEA implies that the evaluations based on it are very sensitive to extreme values and outliers in the data. The “curse” of dimensionality of DEA, due to its nonparametric nature (which, however, offers the great advantage of not requiring a functional specification of the efficient frontier), means that it is necessary to have a lot of observations to reach an acceptable level of precision. For example, DEA efficiency scores are biased and usually overestimated (Banker, 1993). Smith (1997) argues that the extent of the overestimation is dependent on the sample size and the complexity of the production process (as indicated by the number of inputs and outputs). Entani, Maeda, and Tanaka (2002) note that the number of DMUs assessed as efficient will increase exponentially as the dimensions of inputs and outputs increase. When there are many input and output variables and only a few DMUs, decision makers may find that most DMUs are efficient.

Those well-known drawbacks of DEA may lower its robustness when conducting performance analysis. However, in this paper, we mainly focus on methodological ideas and show the capability of DEA approach as a grading tool using multi-level frontiers instead of improving the performance of DEA approach.

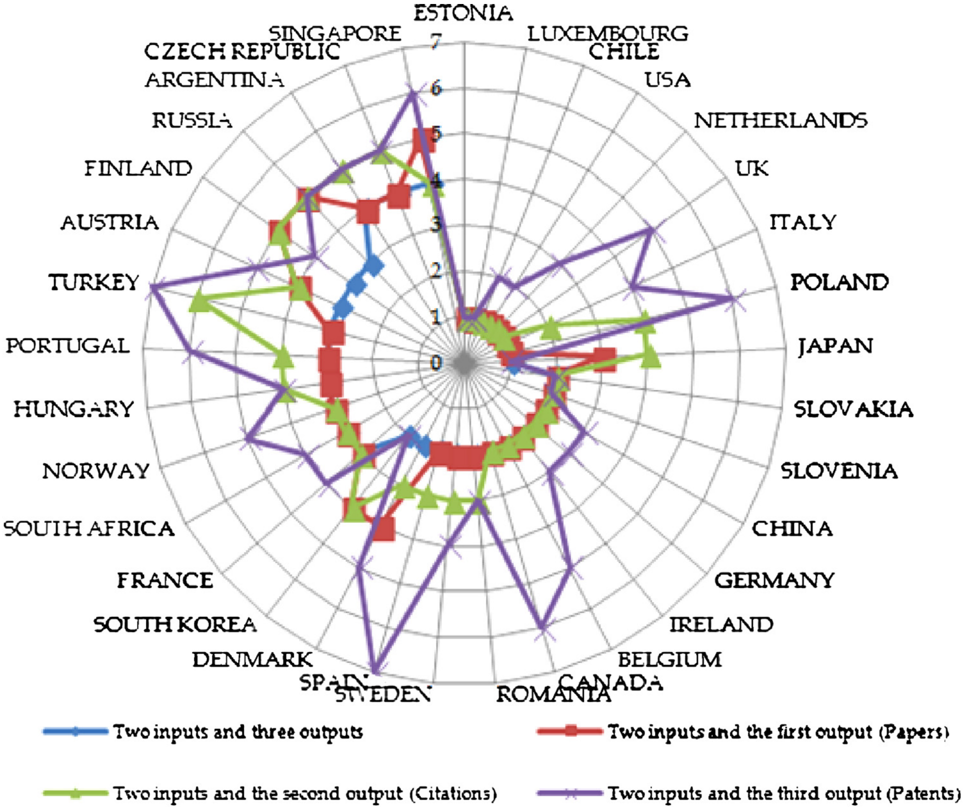

Fig. 7. Visualization of the grades of the countries/territories.

5. Results

5.1. Using BCC–DEA to produce multi-level efficient frontiers and performance targets

In this paper the BCC–DEA model is applied to produce multi-level efficient frontiers, and these were used to decompose the countries/territories of the study into different grades.

The results show that Chile, Estonia, Italy, Japan, Luxembourg, Netherlands, Poland, UK and USA are the first grade countries in the sense of performance level (Fig. 6). Argentina, Czech Republic and Singapore are units with the lowest performance among the 33 countries/territories and belong to the last grade (Grade 4). Table A1 (Appendix A) shows the multi-level performance scores of the countries/territories. We also present the performance improvement chain of these countries/territories in Fig. 6. We first take Argentina as example. Argentina is located on the fourth performance level, so it is not realistic to set a target for Argentina on Grade 1. It is better to find a target on the adjacent Grade 3. Thus we can see that the target of this country is a weighted combination of Hungary and Portugal with weights 0.5453 and 0.4547, respectively. China is located on the second grade of performance and the targets of China are Japan, UK and USA with corresponding weights 0.1651, 0.6354 and 0.1996. Fig. 6 provides a wealth of information on those countries/territories.

5.2. Changes of grades of the countries/territories

We reused the multi-level efficient frontiers in the BCC–DEA model on the 33 countries/territories with two inputs and the first output (Papers), the second output (Citations) and the third output (Patents), respectively, to decompose the countries/territories into different grades. Table 4 reports the grades of the countries/territories. It is clear that some countries (e.g., UK) perform better on Papers and Citations than Patents. However some countries, like Japan, perform much better on Patents than Papers and Citations.

Fig. 7 corresponds to Table 4 and visualizes the grades of the countries/territories when using two inputs and three outputs, two inputs and the first output (Papers), two inputs and the second output (Citations), and two inputs and the third output (Patents).

It is surprising that Chile is rated first grade countries together with S&T-developed countries like USA and UK. For Papers, Citations and Patents as output, respectively, we can verify this result using the ratios Papers to Researcher, Papers to Research Funding, Citations to Researchers, Citations to Research Funding, Patents to Researcher, and Patents to Research

Table 5
Ratios of output indicator values to input indicator values.

No.	Countries/ Territories	Papers/ researcher	Papers/ research funding	Citations/ researcher	Citations/ research funding	Patents/ researcher	Patents/ research funding
1	Argentina	0.1630	1.9258	1.5737	18.5961	0.0044	0.0525
2	Austria	0.3303	1.2509	4.8606	18.4082	0.1214	0.4596
3	Belgium	0.4417	1.9995	6.9381	31.4069	0.1227	0.5555
4	Canada	0.3649	2.2848	5.1757	32.4083	0.0629	0.3941
5	Chile	0.9082	4.7301	8.7100	45.3615	0.0349	0.1816
6	China	0.1088	0.6709	1.0396	6.4104	0.0712	0.4393
7	Czech Republic	0.3088	2.2277	3.1775	22.9213	0.0230	0.1657
8	Denmark	0.3296	1.8270	5.7493	31.8704	0.1009	0.5595
9	Estonia	0.3184	2.5889	4.0851	33.2194	0.0190	0.1546
10	Finland	0.2545	1.3643	3.7493	20.0952	0.1500	0.8039
11	France	0.2720	1.2946	3.7109	17.6651	0.1401	0.6669
12	Germany	0.2810	1.0245	4.1067	14.9713	0.2163	0.7885
13	Hungary	0.2736	2.3461	3.0532	26.1783	0.0270	0.2316
14	Ireland	0.4603	2.2057	6.6901	32.0600	0.1124	0.5385
15	Italy	0.5177	2.1195	6.9713	28.5407	0.1885	0.7717
16	Japan	0.1178	0.5304	1.2477	5.6184	0.4375	1.9702
17	Luxembourg	0.2180	0.8640	2.6711	10.5839	0.3334	1.3213
18	Netherlands	0.5625	2.3841	9.7635	41.3847	0.2560	1.0852
19	Norway	0.3757	2.0447	5.1840	28.2145	0.0831	0.4524
20	Poland	0.3195	3.5144	2.4230	26.6523	0.0294	0.3232
21	Portugal	0.2357	2.3636	2.7944	28.0200	0.0070	0.0702
22	Romania	0.3612	4.0072	2.0734	23.0031	0.0281	0.3123
23	Russia	0.0636	0.8239	0.3479	4.5091	0.0550	0.7128
24	Singapore	0.2929	1.2063	4.4745	18.4295	0.0574	0.2365
25	Slovakia	0.2112	3.7516	1.7158	30.4784	0.0086	0.1525
26	Slovenia	0.4433	2.9038	4.0170	26.3113	0.0568	0.3719
27	South Africa	0.4276	1.7700	4.2447	17.5710	0.0648	0.2681
28	South Korea	0.1551	0.7897	1.4228	7.2445	0.3183	1.6207
29	SPAIN	0.3598	2.3409	4.4232	28.7752	0.0361	0.2348
30	Sweden	0.4252	1.5826	6.7005	24.9419	0.2129	0.7927
31	Turkey	0.3474	2.2806	2.1080	13.8374	0.0127	0.0830
32	UK	0.3824	2.4955	5.8669	38.2838	0.0666	0.4343
33	USA	0.2879	0.8490	4.4277	13.0554	0.1511	0.4455

Funding. From Table 5, we see that Chile performs very well for the ratios of Papers to Researcher, Papers to Research Funding, Citations to Researchers, and Citations to Research Funding. On the contrary, we see that China, Japan, South Korea and Russia have a low performance compared to other countries for the ratios of Papers to Researcher, Papers to Research Funding, Citations to Researchers, and Citations to Research Funding. We believe that a reason for this is that researchers from these countries publish relatively frequently in domestic journals that are not covered by the WoS. However Japan and South Korea perform very well for the ratios of Patents to Research Funding, which means these two countries pay a lot of attention to technological innovation.

5.3. Incorporating prior information into BCC–DEA

In this case, there are two inputs, three outputs and 33 countries/territories. We incorporate prior information, transforming Model (8) into Model (9), in which constraint $u_1 \leq u_2$ is added to ensure that more weight is put on Citations compared to Papers as the prior information:

$$
s.t.\begin{cases}
\max_{u_r,v_i,\mu_0} \sum_{r=1}^{3} u_r y_{r0} + \mu_0 \\
\sum_{r=1}^{3} u_r y_{rj} - \sum_{i=1}^{2} v_i x_{ij} + \mu_0 \leq 0, \quad j = 1, \ldots, 33 \\
\sum_{i=1}^{2} v_i x_{i0} = 1 \\
u_1 \leq u_2 \\
u_r \geq 0, u_2 \geq 0, u_3 \geq 0, v_1 \geq 0, v_2 \geq 0, \mu_0 \quad \text{free}
\end{cases} \tag{9}
$$

We used Model (9) to obtain additional performance grading results. When number of citations is considered more important than the number of papers, Chile, Estonia, Japan, Luxembourg, Netherlands, UK, and USA are labeled Grade 1 in the sense of performance level (Table 6), whereas Turkey on its own forms the poorest performance unit (Grade 5) among the 33 countries/territories. Detailed data is given in Table 6.

Table 6
Different grades of the countries/territories with two inputs and three outputs, when number of citations is considered more important than number of papers (Model (9)).

Grades	Countries/territories
Grade 1	Chile, Estonia, Japan, Luxembourg, Netherlands, UK, USA
Grade 2	Belgium, Canada, China, Denmark, Germany, Ireland, Italy, Slovakia, Slovenia, South Korea, Sweden
Grade 3	Austria, Finland, France, Hungary, Norway, Portugal, Romania, Russia, South Africa, Spain
Grade 4	Argentina, Czech Republic, Poland, Singapore
Grade 5	Turkey

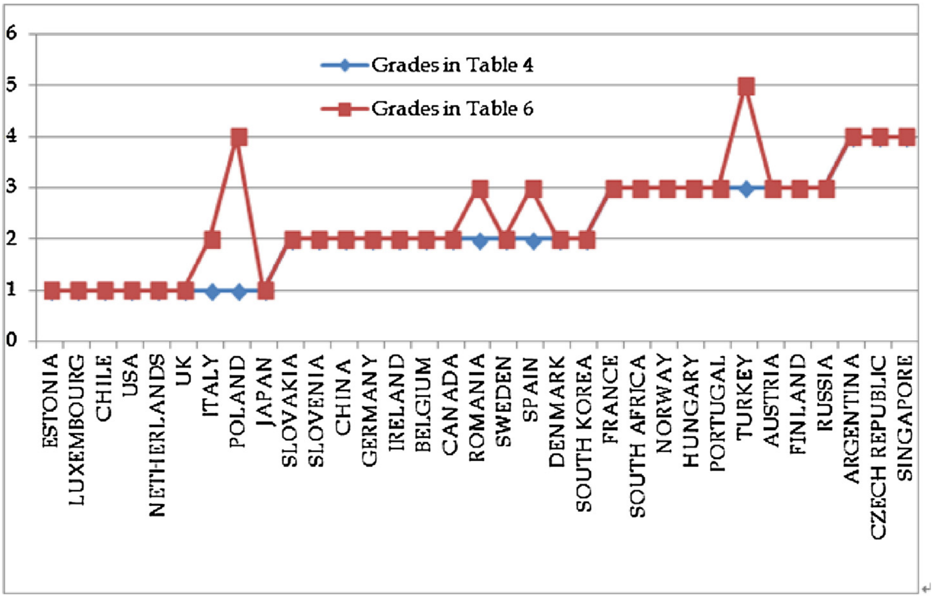

Fig. 8. Grades of the countries/territories in Table 4 (two inputs and three outputs) and Table 6 (if only one line is visible, this means that the two lines coincide).

When we compare the grades of the countries/territories in Table 4 (two inputs and three outputs) and Table 6, we observe, for instance, that the grades of Poland and Turkey drop significantly. This fact shows that the performance of these two countries decrease when we incorporate the prior information that the number of citations is more important than the number of papers. In Fig. 8, the grades of the countries/territories in Tables 4 and 6 are shown graphically.

Also in this case, even though we do not do it in this paper, it is possible to show the performance improvement chain of the countries/territories (cf. Section 5.1).

6. Discussions and conclusions

In this paper we have shown that multi-level frontiers of DEA can be used to classify countries/territories into different grades, reflecting the level of performance of the selected countries/territories. The approach put forward is not restricted to grading countries/territories: it can also be used to grade, for instance, journals and research institutions based on properly selected indicators. Lee and Shin (2014) proposed a new measure of journal performance that captures field-different citation characteristics, a measure that is based on DEA through viewing journal performance from the perspective of the efficiency of a journal's citation generation. It is also possible to measure the performance of journals from an output perspective only. In other words, in case of no explicit inputs, e.g., when journals should be graded using only output indicators, we can assume that there is single constant input, which is equal to unity for all observations (e.g., Yang, Shen, Zhang, & Liu, 2014b).

There are two main advantages of the grading approach proposed in this paper. (1) It is a nonparametric and recursive approach, which needs no a priori information like indicator weights and threshold values for different grades. (2) The observations within the same grade are indifferent in the sense of performance level. The main disadvantage of the approach is that in some cases there are too few indicators (single input and single output). Under such circumstances, it could happen that each grade contains exactly one observation (in our case, exactly one DMU). Thus, the approach is more suitable for grading observations with multiple input and output indicators.

We need to point out that our empirical analysis has some critical limitations. First, we do not take into account how the gross domestic expenditures on R&D are distributed over institutional type within the countries/territories. The proportion of gross domestic expenditures on R&D in public research institutions varies among countries/territories. Moreover, public research institutions publish considerably more than the private sector. If the proportion of gross domestic expenditures on R&D in public research institutions of country *A* is, say, twice that of the corresponding proportion of country *B*, then *A* will perform better then *B* with respect to publication output, keeping all other things constant.

Second, we do not take into account how the gross domestic expenditures on R&D are distributed over disciplines within the countries/territories. This matters because publication intensity varies across disciplines. It is, for instance, much higher in the sciences than in the humanities. Further, the proportion of gross domestic expenditures on R&D spent among disciplines probably varies across countries/territories. If country *A* invest proportionally more than country *B* in the sciences, then *A* will perform better then *B* with respect to publication output, given that all other things being held constant. Moreover, the citations we deal with are not field-normalized. Therefore, *A* will perform better then *B* also with respect to citations if the constant condition is satisfied.

A circumstance that might be considered as a third limitation is that our two input indicators are linearly correlated, since a substantial share of the gross domestic expenditures on R&D concerns labor. It might be tempting, with the purpose to increase discrimination, to omit input or output variables that are linearly correlated. However, as is shown in Amirteimooria, Despotis, and Kordrostami (2014), this may influence the performance analysis. These authors propose an iterative variable reduction approach in which each two highly correlated variables are aggregated in each iteration. We might take this approach into consideration in future research.

In view of the limitations discussed above, we believe that the accuracy level of the empirical analysis of our study is insufficient for actual policy-level decision making. However, the contribution of this work is mainly in proposing new methodological ideas. We focus on showing the capability of the DEA approach as a grading tool using multi-level frontiers, and do not try to increase the accuracy of the empirical analysis of DEA applications.

For future research, we would like to explore more features of the multi-level DEA frontiers regarding weight restrictions in DEA models with more prior information. There are at least four types of restrictions on the weights of input and output variables (e.g., Allen et al., 1997), and the efficient frontiers will vary accordingly and show different properties. Finally, we recall that our grading approach can easily be applied to the classification of scientific journals, research institutions, etc.

Acknowledgements

We would like to acknowledge the support of the National Natural Science Foundation of China (NSFC, No. 71201158; No. L1422061). We thank Ludo Waltman, editor of the Journal of Informetrics, and two anonymous referees for valuable comments which helped us to significantly improve the manuscript.

Appendix A.

Table A1.

Table A1
Performance scores in BCC–DEA model (2 inputs and 3 outputs).

Countries/territories	Performance scores in input-based BCC–DEA model (Model (6))			
	Grade 1	Grade 2	Grade 3	Grade 4
DMU1-Argentina	0.4858	0.6351	0.8173	1
DMU 2-Austria	0.5432	0.7508	1	–
DMU 3-Belgium	0.7641	1	–	–
DMU 4-Canada	0.9049	1	–	–
DMU 5-Chile	1	–	–	–
DMU 6-China	0.6100	1	–	–
DMU 7-Czech Republic	0.6183	0.7984	0.9928	1
DMU 8-Denmark	0.7596	1	–	–
DMU 9-Estonia	1	–	–	–
DMU 10-Finland	0.6518	0.8659	1	–
DMU 11-France	0.7102	0.9054	1	–
DMU 12-Germany	0.9522	1	–	–
DMU 13-Hungary	0.5906	0.8512	1	–
DMU 14-Ireland	0.7509	1	–	–
DMU 15-Italy	1	–	–	–
DMU 16-Japan	1	–	–	–
DMU 17-Luxembourg	1	–	–	–
DMU 18-Netherlands	1	–	–	–
DMU 19-Norway	0.6670	0.8997	1	–
DMU 20-Poland	1	–	–	–
DMU 21-Portugal	0.7032	0.8931	1	–
DMU 22-Romania	0.9642	1	–	–
DMU 23-Russia	0.4968	0.5980	1	–

Table A1 (*Continued*)

Countries/territories	Performance scores in input-based BCC–DEA model (Model (6))			
	Grade 1	Grade 2	Grade 3	Grade 4
DMU 24-Singapore	0.4746	0.6576	0.8693	1
DMU 25-Slovakia	0.9641	1	–	–
DMU 26-Slovenia	0.7783	1	–	–
DMU 27-South Africa	0.6015	0.9416	1	–
DMU 28-South Korea	0.8983	1	–	–
DMU 29-Spain	0.9039	1	–	–
DMU 30-Sweden	0.7983	1	–	–
DMU 31-Turkey	0.8105	0.9381	1	–
DMU 32-UK	1	–	–	–
DMU 33-USA	1	–	–	–

Note:—Means that the DMU in this grade has been removed from the observations.

References

Abbott, M., & Doucouliagos, C. (2003). The efficiency of Australian universities: A data envelopment analysis. *Economic of Education Review*, *22*(1), 89–97.
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.
Amirteimooria, A., Despotis, D. K., & Kordrostami, S. (2014). Variables reduction in data envelopment analysis. *Optimization*, *63*(5), 735–745.
Aristovnik, A. (2012). The relative efficiency of education and R&D expenditures in the new EU member states. *Journal of Business Economics and Management*, *13*(5), 832–848.
Avkiran, N. K. (2001). Investigating technical and scale efficiencies of Australian universities through data envelopment analysis. *Socio-Economic Planning Sciences*, *35*(1), 57–80.
Bandyopadhyay, A. (2013). *Ranking of Business School Journals: A Rating Guide for Researchers*. Retrieved from: ⟨http://mpra.ub.uni-muenchen.de/49608/1/MPRA_paper_49608.pdf⟩ (Aug. 23, 2014)
Banker, R. D. (1993). Maximum likelihood, consistency and data envelopment analysis: A statistical foundation. *Management Science*, *39*(10), 1265–1273.
Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, *30*(9), 1078–1092.
Banker, R. D., Cooper, W. W., Seiford, L. M., Thrall, R. M., & Zhu, J. (2004). Returns to scale in different DEA models. *European Journal of Operational Research*, *154*, 345–362.
Bonaccorsi, A., & Daraio, C. (2005). Exploring size and agglomeration effects on public research productivity. *Scientometrics*, *63*(1), 87–120.
Brandt, T., & Schubert, T. (2013). Is the university model an organizational necessity? Scale and agglomeration effects in science. *Scientometrics*, *94*(2), 541–565.
Carrington, R., Coelli, T., & Rao, P. D. S. (2005). The performance of Australian universities: Conceptual issues and preliminary results. *Economic Papers-Economic Society of Australia*, *24*(2), 145–163.
CAS. (2006). *Report on comprehensive quality evaluation of institutes in Chinese academy of sciences (CAS)*. Retrieved from ⟨http://www.bps.cas.cn/kjpj/xgzl/200905/t20090527_232560.html⟩ (Aug. 23 2014).
Charnes, A., Cooper, W. W., & Rhodes, E. L. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429–444.
Chen, C. P., Hu, J. L., & Yang, C. H. (2011). R&D efficiency of multiple innovative outputs: The role of the national innovation system. *Innovation: Management, Policy & Practice*, *13*, 341–360.
Cook, W. D., & Seiford, L. M. (2009). Data envelopment analysis (DEA)—Thirty years on. *European Journal of Operational Research*, *192*, 1–17.
Cook, W. D., Tone, K., & Zhu, J. (2014). Data envelopment analysis: Prior to choosing a model. *Omega*, *44*, 1–4.
Cooper, W. W., Seiford, L. M., & Tone, K. (2007). *Data envelopment analysis: A comprehensive text with models, applications, references and DEA-solver software* (2nd ed.). New York, NY: Springer.
Daraio, C., Bonaccorsi, A., & Simar, L. (2015). Efficiency and economies of scale and specialization in European universities: A directional distance approach. *Journal of Informetrics*, *9*(3), 430–448.
Entani, T., Maeda, Y., & Tanaka, H. (2002). Dual models of interval DEA and its extension to interval data. *European Journal of Operational Research*, *136*(1), 32–45.
Flegg, A. T., Allen, D. O., Field, K., & Thurlow, T. W. (2004). Measuring the efficiency of British universities: A multi-period data envelopment analysis. *Education Economics*, *12*(3), 231–249.
Glänzel, W. (2011). The application of characteristic scores and scales to the evaluation and ranking of scientific journals. *Journal of Information Science*, *37*(1), 40–48.
Grosskopf, S., Hayes, K. J., & Taylor, L. L. (2014). Applied efficiency analysis in education. *Economics and Business Letters*, *3*(1), 19–26.
Harvey, C., Kelly, A., Morris, H., & Rowlinson, M. (2010). *Academic journal quality guide—Version 4*. London: Association of Business Schools.
Harvey, C., Morris, H., & Kelly, A. (2007a). *Academic journal quality guide: Context, purpose and methodology*. London: Association of Business Schools.
Harvey, C., Morris, H., & Kelly, A. (2007b). *Academic journal quality guide*. London: Association of Business Schools.
Harvey, C., Morris, H., & Kelly, A. (2008). *Academic journal quality guide version 2: Context, purpose and methodology*. London: Association of Business Schools.
Hatzichronoglou, T. (1997). Revision of the high-technology sector and product classification. In *OECD science, technology and industry working papers, no. 1997/02*. http://dx.doi.org/10.1787/134337307632
Johnes, G., & Johnes, J. (1992). Apples and oranges: The aggregation problem in publications analysis. *Scientometrics*, *25*(2), 353–365.
Kao, C., & Lin, Y. C. (1999). Comparing university libraries of different university size. *Libri*, *49*(3), 150–158.
Kempkes, G., & Pohl, C. (2010). The efficiency of German universities—Some evidence from non-parametric and parametric methods. *Applied Economics*, *42*(16), 2063–2079.
Lee, H., & Shin, J. (2014). Measuring journal performance for multidisciplinary research: An efficiency perspective. *Journal of Informetrics*, *8*(1), 77–88.
Lee, H. Y., & Park, Y. T. (2005). An international comparison of R&D efficiency: DEA approach. *Asian Journal of Technology Innovation*, *13*(2), 207–222.
Lee, H. Y., Park, Y. T., & Choi, H. (2009). Comparative evaluation of performance of national R&D programs with heterogeneous objectives: A DEA approach. *European Journal of Operational Research*, *196*(3), 847–855.
Rousseau, S., & Rousseau, R. (1997). Data envelopment analysis as a tool for constructing scientometric indicators. *Scientometrics*, *40*(1), 45–56.
Rousseau, S., & Rousseau, R. (1998). The scientific wealth of European nations: Taking effectiveness into account. *Scientometrics*, *42*(1), 75–87.
Roy, S., & Nagpaul, P. S. (2001). A quantitative evaluation of relative efficiencies of research and development laboratories: A data envelopment analysis approach. In M. Davis, & C. S. Wilson (Eds.), *Proceedings of the eighth international conference on scientometrics & informetrics* (pp. 629–638). Sydney (Australia): BIRG.
Schubert, T. (2014). Are there scale economies in scientific production? On the topic of locally increasing returns to scale. *Scientometrics*, *99*(2), 393–408.

Seiford, L. M., & Zhu, J. (2003). Context-dependent data envelopment analysis—Measuring attractiveness and progress. *Omega—The International Journal of Management Science*, *31*, 397–408.

Sharma, S., & Thomas, V. J. (2008). Inter-country R&D efficiency analysis: An application of data envelopment analysis. *Scientometrics*, *76*(3), 483–501.

Shim, W., & Kantor, P. B. (1998). A novel economic approach to the evaluation of academic research libraries. *Proceedings of the ASIS Annual Meeting*, *35*, 400–410.

Smith, R. (1996). Peer review: A flawed process in the heart of science and journals. *Journal of the Royal Society of Medicine*, *99*(4), 178–182.

Smith, P. (1997). Model misspecification in data envelopment analysis. *Annals of Operations Research*, *73*, 233–252.

Strayer, J. K. (1989). *Linear programming and applications*. New York: Springer-Verlag.

Sueyoshi, T., & Goto, M. (2013). A use of DEA–DA to measure importance of R&D expenditure in Japanese information technology industry. *Decision Support Systems*, *54*, 941–952.

Wang, E. C., & Huang, W. C. (2007). Relative efficiency of R&D activities: A cross-country study accounting for environmental factors in the DEA approach. *Research Policy*, *36*(2), 260–273.

Wolszczak-Derlacz, J., & Parteka, A. (2011). Efficiency of European public higher education institutions: A two-stage multicountry approach. *Scientometrics*, *89*(3), 887–917.

Worthington, A. C. (2001). An empirical survey of frontier efficiency measurement techniques in education. *Education Economics*, *9*(3), 245–268.

Worthington, A. C., & Lee, B. L. (2008). Efficiency, technology and productivity change in Australian universities, 1998–2003. *Economics of Education Review*, *27*(3), 285–298.

Yang, G. L., Ahlgren, P., Yang, L. Y., Rousseau, R., & Ding, J. L. (2015). Grading countries/territories using DEA frontiers. In A. Salah, Y. Tonta, A. A. A. Salah, C. Sugimoto, & U. Al (Eds.), *Proceedings of ISSI 2015* , ISBN 978-975-518-381-7 (pp. 436–447). Istanbul: Boğaziçi University Printhouse.

Yang, G. L., Yang, L. Y., Liu, W. B., Li, X. X., & Fan, C. L. (2013). Directional returns to scale of biological institutes in Chinese Academy of Sciences. In J. Gorraiz, E. Schiebel, C. Gumpenberger, M. Hörlesberger, & H. Moed (Eds.), *Proceedings of ISSI 2013* (pp. 551–566). Vienna: Austrian Institute of Technology (AIT).

Yang, G. L., Rousseau, R., Yang, L. Y., & Liu, W. B. (2014). A study on directional returns to scale. *Journal of Informetrics*, *8*(3), 628–641.

Yang, G. L., Shen, W. F., Zhang, D. Q., & Liu, W. B. (2014). Extended utility and DEA models without explicit input. *Journal of the Operational Research Society*, *65*, 1212–1220.

Yang, L. Y., Zhou, Q. J., & Yue, T. (2013). China's science: The overall development and the balance of disciplinary structure—Statistics and analysis of SCI-indexed papers in 2012. *Science Focus*, *8*(1), 23–50 (In Chinese).

Yang, H.-H., & Chang, C.-Y. (2009). Using DEA window analysis to measure efficiencies of Taiwan's integrated telecommunication firms. *Telecommunications Policy*, *33*(1–2), 98–108.

Zhong, W., Yuan, W., Li, S. X., & Huang, Z. M. (2011). The performance evaluation of regional R&D investments in China: An application of DEA based on the first official China economic census data. *Omega—The International Journal of Management Science*, *39*(4), 447–455.