ISSN 1748-7595 (Online)

Kent Business School

Working Paper Series

Multi-level DEA Approach in Research Evaluation

Wenbin Liu
Kent Business School

Wei Meng
Institute of Policy and Management
Chinese Academy of Sciences

Daqun Zhang
Institute of Policy and Management
Chinese Academy of Sciences

University of Kent

Working Paper No. 140
May 2007

Multi-level DEA Approach in Research Evaluation

Wei Meng and Daqun Zhang
Institute of Policy and Management, Chinese Academy of Sciences, 100080, Beijing, China
Wenbin Liu
Kent Business School, University of Kent, CT2 7PE, Canterbury, Kent, UK

Abstract: It is well known that the discrimination power of DEA models will be diminishing if too many inputs or outputs are used. It is a dilemma if the decision makers want to select comprehensive indicators to present a relatively holistic evaluation using DEA. In this work we show that by utilizing hierarchical structures of input-output data DEA can handle quite large numbers of inputs and outputs. We present two approaches in a pilot evaluation of 15 institutes for basic research in Chinese Academy of Sciences using DEA models.

Key words: hierarchical structures, discrimination power, DEA, research evaluation

Introduction

Nowadays, performance evaluation and benchmarking become routine practice in performance management. It has also been well recognised that a single indicator may not be sufficient for performance management, especially for performance evaluation of research institutions, which normally have multi-dimensionalities of research activities. It is now a usual practice to set or select a set of performance indicators in evaluations of research institutions.

For evaluation of Decision Making Units (DMUs) with multiple-inputs and outputs in public sector, Data Envelopment Analysis (DEA) is now one of the most widely accepted methods to measure relative efficiency or productivity. However, it is well known that the discrimination power of DEA models will be diminishing if too many inputs or outputs are used. It is a dilemma if the decision makers (DMs) want to select more indicators to present a relatively holistic evaluation using DEA. This is especially the case in evaluation of large research institutes like those in Chinese Academy of Sciences (CAS), where usually many outputs are measured in evaluation to produce relatively comprehensive profiles of these institutes, see Meng (2006).

Intuitively, people may wish to use some statistical techniques to reduce numbers of indicators in order to improve DEA discrimination. In practical applications, there have quite a few papers proposed different techniques on indicators reduction or aggregation, such as dropping highly correlation indicators, see Kao, Chang and Hwang (1993), Zhang and Bartels (1998), Jenkins and Anderson (2003), Farzipoor Saen, Memariani and Hosseinzadeh Lotfi (2005), or selecting principle components by principle component analysis (PCA), see Adler and Golany (2001), or aggregating indicators by analysis hierarchic process (AHP), see Shang and Sueyoshi (1995), Cai and Wu (2001), Korhonen *et al.* (2001), and Yang and Kuo (2003) *etc*.

However, being an extreme point approach, the standard DEA models are sensitive to indicator set changes, and even removal of a highly correlated output (or inputs) can much change the evaluation results, see Dyson *et al.* (2001). Furthermore removal of highly correlated data may not be rational in research evaluations, where it is well accepted that research may have many outputs and their consequences like papers, citations of publications, awards, and invited talks, *etc*., which are complementary but often highly correlated. Often the DMs wish to include many such correlated indicators in order to reflect the complexity of the research activities more completely. It is difficult to justify partial removals of the indicators just because of data correlations.

It has been observed that in research evaluation, often these indicators can be grouped hierarchically, where different weights can be relatively easily assigned to reflect their relative importance **within the groups, while no such substitutions can be easily decided between these groups** so they are best considered to be no-substitutable. In this paper, we carry out a pilot study on DEA productivity evaluation of 15 institutes for basic research in CAS by exploring multi-level data structures. The main purpose of this investigation is to explore the possibility of using DEA for efficiency evaluation of CAS, where a large numbers of indicators are grouped hierarchically so that the standard DEA models have not been able to be applied.

Inputs and outputs used in CAS research institute evaluation

One of main objectives of CAS is "**do better basic research"**. Actually, CAS is a major player in basic research of China. Following in-depth reformation of Knowledge Innovation Program (KIP) of the CAS, which was launched in 1998, research quantity and quality of basic research have been steadily increasing. In evaluation of sustainability in Comprehensive Evaluation System (CES) in 2002, research outcomes were measured from three aspects: goal achievement, quantitative measurements, and social and economic contributions, see Li (2005). Goal achievement was evaluated by peer review based on the pre-signed short-term (3 years) research contracts between the CAS administration and its research institutes. Quantitative measurement was based on three indicators as: high quality publications, the number of publications in top research journals in subject disciplines; invited talks in top international conferences, important national and international awards. Then patents commercialisation, joint companies, rewarded invention patents, significant consultant reports and national standards setting-up were selected as the indicators to reflect social and economic contribution of basic research. With these selected indicators and assigned weights, weighted sum of sub-scores of various indexes and volume data was used as the overall performance scores in CES 2002, although rationality of the weights selection has been questioned since the CAS evaluation system started. These provided us the initial motivations to apply DEA approach on performance evaluation of the research institutes in CAS, especially on research productivity evaluation. Since DEA allows each institute to exhibit its best performance with full flexibility of weights selection, thus the problem of weight selection could be dealt with by using DEA.

In DEA applications, inputs and outputs need to be decided in advance. For research evaluation of basic research institutes, usually the inputs are quite straightforward to decide. The number of research staff, advanced research equipment, and total research expenditures are major hard research inputs for research activities. However, the available data of research equipment from the Statistical Yearbook of the CAS only classified all the equipment into two categories: purchased before 1990 or after 1990. And only total original values were provided for each institute. Because the detailed data were not available, we had to omit the equipment here. Thus two inputs are selected: one is the numbers of researchers, which are counted using total permanent research staff plus post-doctors. The second input is research expenditures. Because of special Chinese culture, pensions of retired staff are still provided by the institutes. Thus data of research expenditures need to exclude this part. Besides, there are also some soft inputs that can benefit research outputs. However, data of this type are also not available.

Research outputs are numerous depending on different stakeholders' view. Nowadays, direct research outputs, research competitiveness and scientist cultivation are the three main evaluation aspects on performance evaluation of institutes for basic research in the CAS. The following represents a view from the level of Bureau of Basic Sciences on the most important top performance indicators of the 15 institutes for basic research in the CAS are research outputs, external research funding, and scientist cultivation respectively, see Meng *et al.* (2005). These indicators are also frequently used on performance evaluation of basic research, see Kaukonen (1997), Glänzel, Schubert and Braun (2002), King (2004), Gracia and Sanz-Menéndez (2005), Meng, Hu and Liu (2006) *etc*.

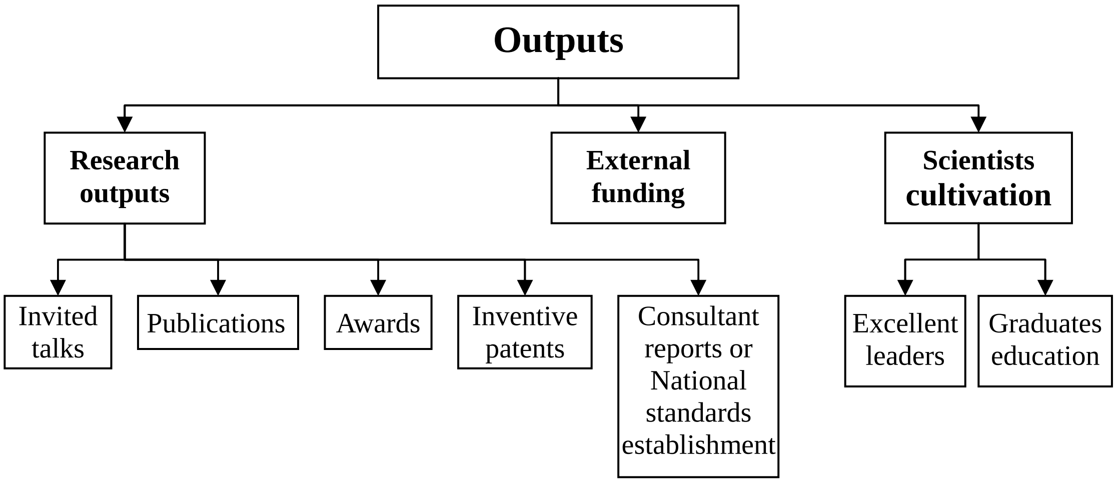

Fig 1 Research outcomes with hierarchical structure

Meng *et al.* (2005) presented a questionnaire analysis, where AHP has been used to judge relative importance for some research outputs of 15 CAS research institutes of basic research, where the selected research outputs can be further constructed in three levels, as Figure 1 shown. Direct research outputs, external research funding obtained to measure research competitiveness, referred to as external funding and scientists cultivation are on the top level. On the second level, research outputs can be further decomposed into five sub-indicators, as invited talks, publications, awards, inventive patents and consultant reports or national standards establishment. Scientist cultivation includes excellent research leaders fostering and postgraduates educations. The relative importance of the selected sub-indicators obtained via AHP as shown in Table 1, more details can be found in Meng *et al.* (2005).

Table1: Weights of 5 sub-indicators of research outputs based on AHP

eigenvalue		publications	awards	Invited talks	Invention patents	Cons. Rep. Stand. Estab.
Geometry Mean A_{geo}	5.8996	0.23593	0.42606	0.36661	0.2195	0.17549
Percent% (weights)		16.57	29.93	25.75	15.42	12.33

This information will be used as a base for formulation of multi-level DEA approaches.

Since we only have 15 DMUs, it seems impossible to directly handle these indicators using DEA. In the next section we first adopt the weights generated from AHP to aggregate related indicators from bottom to up, and then apply DEA at the second level. This approach is related to the approaches of data transformation as discussed in Liu *et al.* (2006) and Thanassoulis, Portela and Allen (2004). Then we build new DEA models to incorporate these hierarchical structures directly in the section afterwards.

DEA with data transformation

In practical applications, AHP and DEA have been combined to tackle many complex problems in performance evaluation. For instance in order to apply DEA, qualitative indicators need to be quantified firstly. AHP has the advantage in dealing with subjective factors and summarising them into a set of numerical indicators. Then DEA can be applied, see Shang and Sueyoshi (1995), and Yang and Kuo (2003). On the other hand, some papers applied DEA firstly generate scores to form a pair-wise comparison matrix, and then AHP to generate weights of units from the matrix, see Sinuany-Stern, Abraham and Yossi (2000), Ramanathan (2006), and Liu and Hai (2005). AHP also can be used to aggregate indicators based on the weights that are derived from the AHP approach, see Cai and Wu (2001), and Korhonen, Tarnio and Wallenius (2001), and we here adopt this approach.

In this case, the five sub-indicators are aggregated into research outputs according to the weights obtained in the table. For the third output– scientist cultivation, the two sub-indicators are used: 1) the number of excellent research leaders is regarded as an indicator to reflect research sustainability and reputation implicitly; 2) graduate education is to reflect sustainability and social accountability of basic research as well as research reputation. Thus excellent research leaders and graduates enrolment are selected to reflect scientist cultivation of institutes for basic research. However the relative importance of these two sub-indicators was not available. In this case we regard these two indicators to be equally important as a pilot study.

Thus we have two inputs – research staff and expenditure, and three aggregated outputs – direct research outputs, research competitiveness (via external research funding), and scientists cultivation. Then the normalised data is shown in Table 2. The formula to normalise these indexes is $\bar{y}_{ij} = \frac{y_{ij}}{\underset{i}{Max}\, y_{ij}} \times 100$. The purpose to normalise indexes is to remove scale differences in these weighted sums.

Table 2: Normalised data with aggregated output

DMUs	Staff	Research. Expenditure.	Research outputs	External funding	Scientist Cultivation
Unit 1	34.39	44.61	49.89	47.96	**100.00**
Unit 2	37.83	59.53	62.60	94.60	70.47
Unit 3	6.15	9.80	6.27	7.79	22.36
Unit 4	**100.00**	69.06	11.90	66.54	72.06
Unit 5	22.44	14.09	4.43	18.71	18.45
Unit 6	74.93	**100.00**	14.92	**100.00**	39.09
Unit 7	43.53	39.08	4.38	44.47	18.03
Unit 8	44.62	30.42	3.25	17.28	18.39
Unit 9	17.92	17.22	3.13	21.65	10.09
Unit 10	21.99	24.27	27.96	28.11	4.67
Unit 11	50.05	46.26	10.19	92.11	19.82
Unit 12	31.40	36.88	11.69	42.87	23.29
Unit 13	40.27	58.31	**100.00**	84.68	73.08
Unit 14	23.53	20.51	41.44	46.04	15.73
Unit 15	27.51	44.29	17.34	94.94	44.80

Now let us consider DEA models. In this section we will firstly apply the BCC models on the aggregated data in Table 2.

$$\max \quad \theta + \varepsilon\left(\sum_{i=1}^{2} s_i^- + \sum_{r=1}^{3} s_r^+\right)$$

$$\text{subject to: } \sum_{j=1}^{15} x_{ij}\lambda_j + s_i^- = x_{i0},$$

$$\sum_{j=1}^{15} y_{rj}\lambda_j - s_r^+ = \theta\, y_{r0}, \qquad (1)$$

$$\sum_{j=1}^{15} \lambda_j = 1,$$

$$\theta \geq 1,\ s_i^-,\ s_r^+,\ \lambda_j \geq 0,\ i = 1,2,\ r = 1,2,3,\ j = 1,..,15.$$

In Model 1, Pareto preference and radial measurement are used to confirm that three outputs on the top level are regarded as equally important and no-substitutable. The results are shown in the second column of Table 3, named as Model 1. The third column presents ranking order based DMUs' efficiency score. Furthermore, if evaluators do not prefer the radial measurement, then Model 1 can be further extended with Russell measurement, as Model 2 presented. The formulas $\theta_r \geq 1, r = 1, 2, 3$ in Model 2 confirm that Pareto preference is still used. The results are listed at the fourth column in Table 3, followed by its ranking order for each DMU.

$$\begin{aligned}
\max \quad & \frac{1}{3}\sum_{r=1}^{3}\theta_r + \varepsilon\sum_{i=1}^{2}s_i^- \\
\text{subject to:} \quad & \sum_{j=1}^{15}x_{ij}\lambda_j + s_i^- = x_{i0}, \\
& \sum_{j=1}^{15}y_{rj}\lambda_j = \theta_r\ y_{r0}, \\
& \sum_{j=1}^{15}\lambda_j = 1, \\
& \theta_r \geq 1,\ r = 1,\ 2,\ 3, \\
& s_i^-, \lambda_j \geq 0, i = 1, 2,\ j = 1,..,15.
\end{aligned} \tag{2}$$

On the other hand, if all the aggregated three outputs are regarded as equally substitutable, and the DMs or evaluators prefer to compare their performance in the average sense, then constraint $\theta_r \geq 1, r = 1, 2, 3$ in Model 2 needs to be replaced by $\sum_{r=1}^{3}\theta_r \geq 3,\ \theta_r \geq 0$, which implies that the three outputs can be substituted, and the average level is measured in the objective function, seeing Liu *et al.* (2006) for the theoretical explanation. The efficiency scores are shown in the sixth column, named as Model 3 in Table 3.

$$\begin{aligned}
\max \quad & \frac{1}{3}\sum_{r=1}^{3}\theta_r + \varepsilon\sum_{i=1}^{2}s_i^- \\
\text{subject to:} \quad & \sum_{j=1}^{15}x_{ij}\lambda_j + s_i^- = x_{i0} \\
& \sum_{j=1}^{15}y_{rj}\lambda_j = \theta_r\ y_{r0} \\
& \sum_{j=1}^{15}\lambda_j = 1, \sum_{\text{r}=1}^{3}\theta_r \geq 3, \\
& s_i^-,\ \theta_r,\ \lambda_j \geq 0,\ i = 1,2,\ r = 1,2,3, j = 1,..,15.
\end{aligned} \tag{3}$$

Table 3: Efficiency scores based on DEA models with aggregated indicators

DMUs	**Model 1**	**Rank M1**	**Model 2**	**Rank M2**	**Model 3**	**Rank M3**
Unit 3	100.00	1	100.00	1	100.00	1
Unit 13	100.00	1	100.00	1	100.00	1
Unit 2	100.00	1	100.00	1	90.55	3
Unit 1	100.00	1	100.00	1	83.90	4
Unit 14	100.00	1	100.00	1	77.86	5

Unit 15	100.00	1	100.00	1	54.14	6
Unit 5	86.19	10	43.46	9	43.46	7
Unit 12	56.91	13	33.63	10	33.63	8
Unit 6	100.00	1	100.00	1	31.86	9
Unit 4	87.59	9	28.07	11	28.07	10
Unit 11	96.84	8	45.05	8	26.02	11
Unit 9	63.66	12	23.05	13	23.05	12
Unit 10	65.70	11	23.71	12	22.31	13
Unit 7	52.80	14	15.05	14	15.05	14
Unit 8	36.56	15	13.38	15	13.38	15

Table 3 uses the rank of Rank M3 in sorting data. Comparing the results between Model 1 and Model 2, all the efficient DMUs are same, while the ranking orders of inefficient DMUs are different. The difference seems to come from that of measurements: radial or Russell. Taking Unit 12 as an example, it is ranked at 13th according to Model 1, but moves up to 10th based by Model 2. Measurements of components of Unit 12 by Model 2 are $\theta_1 = 1.46$, $\theta_2 = 5.71$ and $\theta_3 = 1.74$, while the ranking order of Unit 4 drops down to 11th by Model 2 from 9th by Model 1. Measurements of each component of Unit 4 in Model 2 are $\theta_1 = 1.27$, $\theta_2 = 8.4$ and $\theta_3 = 1.01$, whose average is worse than that of Unit 12. From this point of view, DEA models with Pareto preference but Russell measurement can provide more accurate measurement for those inefficient DMUs.

If average preference is preferred, only Unit 3 and 13 are still efficient, and all the rest become inefficient, as Model 3 column shown in Table 3. The difference comes from the difference value judgment (or preference), as Model 3 adopts the average preference, while Pareto preference is used in Model 2. Taking Unit 2 as example, its individual measurements of each output in Model 3 are $\theta_1 = 0.84$, $\theta_2 = 1.49$ and $\theta_3 = 0.99$. These imply that the external research funding obtained of Unit 2 is far behind than its peers (Unit 3 and 3) although the rest two outputs are compatible. One of the key differences is that now the individual measurements can be smaller than one as long as their average is grater or equal one.

Multi-level DEA models

In this section, we will not directly aggregate the indicators by transforming data, but develop suitable DEA models to reflect the hierarchy structures.

Still taking the same case used in above section as example, if all the eight indicators shown in Figure 1 are all preferred by the DMs in order to present a relatively comprehensive evaluation, and if all the indicators, shown in Table 4, are regarded equally important and no substitutable, then the BCC modes should be used, and variable return to scale is assumed. The results are presented in the second column in Table 5. There are 11 DMUs classified as efficient by the BCC model. Obviously, the standard DEA model is not very useful because too many indicators are selected and the hierarchical structure is not used. Even so, as we have discussed, full flexibility on weights selection in the standard DEA models has its unique advantages on judging inefficient DMUs. Even in this case, Unit 7 and 8 are still ranked at the worst. Therefore, there should be no excuse on possible biases on weights selection.

On the other hand, indicators may not be equally important and no substitutable in practical applications. For instance, if the five sub-indicators of research outputs (A_1-A_5) and two sub-indicators of scientists cultivation (C_1-C_2) are allowed to be substituted according to the relative importance within the categories, but the three categories are considered to be equally important and non-substitutable, then a new DEA model can be built to directly reflect the hierarchical structure and incorporate the value judgment of DMs. Some of these indicators such as awards have upper bounds, and thus variable return to scale is assumed. Then we can build a DEA model with two levels, where different sub-systems have different preferences. Pareto preference is assumed for the three outputs on the second level, as Model 4 presented.

Table 4: Normalized 8 output-indicators

Units	Publications scalar (A_1)	Invited talks (A_2)	Awards (A_3)	Invention patents (A_4)	Report (A_5)	External funding (B)	Excellent Leaders (C1)	Graduates Education (C2)
Unit 1	31.01	100.00	0.00	0.00	0.00	47.96	100.00	60.12
Unit 2	93.88	0.00	50.00	53.44	0.00	94.60	57.69	55.14
Unit 3	23.40	0.00	0.00	0.00	0.00	7.79	26.92	8.88
Unit 4	33.80	0.00	0.00	11.45	0.00	66.54	15.38	100.00
Unit 5	12.27	0.00	0.00	4.58	0.00	18.71	3.85	25.70
Unit 6	32.86	14.29	0.00	0.76	0.00	100.00	26.92	35.67
Unit 7	8.55	0.00	0.00	8.40	0.00	44.47	7.69	21.18
Unit 8	12.15	0.00	0.00	0.00	0.00	17.28	11.54	17.91
Unit 9	11.72	0.00	0.00	0.00	0.00	21.65	3.85	12.31

Unit 10	7.12	14.29	0.00	0.76	100.00	28.11	0.00	7.48
Unit 11	35.26	0.00	0.00	3.05	0.00	92.11	15.38	16.36
Unit 12	14.24	0.00	0.00	18.32	16.67	42.87	7.69	29.60
Unit 13	100.00	0.00	100.00	100.00	0.00	84.68	34.62	82.40
Unit 14	30.43	0.00	50.00	36.64	0.00	46.04	3.85	21.34
Unit 15	12.98	0.00	0.00	55.73	0.00	94.94	30.77	40.97

$$\text{Max} \quad \theta+\varepsilon\left(\sum_{i=1}^{2}s_i^- + \sum_{r=1}^{3}s_r^+\right)$$

$$\begin{aligned} \text{Subject to: } & (AX)\lambda + s^- = AX_0 \\ & (BY)\,\lambda - s^+ = \theta(BY_0), \qquad\qquad (4) \\ & \lambda, s^-, s^+ \geq 0, \theta \geq 1 \\ & \mathrm{A}=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix},\ B=\begin{pmatrix} w_1 & w_2 & w_3 & w_4 & w_5 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1\end{pmatrix}\end{aligned}$$

where either the weights can be fixed or allow some variations. In all our computations, we allowed the weights to have 20% variations around their averages. Ideally the weight bounds for those five sub-indicators of research outputs should be decided by some statistical methods like peer review through Delphi or AHP. For instance, Takamura and Tone (2003) incorporated the weights restriction in terms of assurance region type I in assessing the case of sites location, where assurance bounds were calculated by the evaluator's weights (W_{ki}) based on AHP, and the lower and up bounds were defined as $L_{i1,i2} = \min_k \frac{W_{ki1}}{W_{ki2}}$, $U_{i1,i2} = \max_k \frac{W_{ki1}}{W_{ki2}}$. However, in our case, if we adopt this idea to derive weights bounds, its assurance boundary becomes too wide to use, as we had 89 valid questionnaires and scores of each indicator from individual researchers can be between 1 and 9. Thus we still use average weights derived from AHP shown in Table 1as a base, but allow certain flexibility. We will revisit this issue at the end of this section.

As Model 4 presented, apparently 8 indicators are aggregated into three overall outputs. However this model is different from the data aggregations used before or the standard

Cone-Ratio DEA models. For example, here certain flexibility is allowed for the weights, which are not given but to be decided by the mathematical models. Thus the mathematical programming is not linear anymore. The evaluation results are presented in the third column of Table 5.

Table 5: The results of multi-level DEA approaches

DMUs	**BCC score**	**Model 4**	**Rank1 M4**	**Model 5**	**Rank1 M5**	**Model 6**	**Rank M6**
Unit 3	100	100.00	1	100.00	1	100.00	1
Unit 13	100	100.00	1	100.00	1	100.00	1
Unit 2	100	100.00	1	100.00	1	88.56	3
Unit 14	100	100.00	1	100.00	1	74.50	4
Unit 1	100	100.00	1	100.00	1	71.94	5
Unit 15	100	100.00	1	100.00	1	46.21	6
Unit 5	100	86.19	10	37.51	9	37.51	7
Unit 12	71.49	56.90	13	28.85	10	28.85	8
Unit 6	100	100.00	1	100.00	1	25.40	9
Unit 4	100	87.59	9	23.58	11	23.58	10
Unit 10	100	57.54	12	23.04	12	21.87	11
Unit 11	100	96.84	8	38.89	8	21.27	12
Unit 9	66.61	63.66	11	18.06	13	18.06	13
Unit 7	54.6	52.80	14	12.48	14	12.48	14
Unit 8	40.83	36.56	15	10.30	15	10.30	15

Similarly, we can use Pareto preference but Russell measurement. Then we have Model 5, which is written for individual components of the inputs and outputs. The results are shown at Model 5 column followed by its ranking orders in Table 5.

$$
\begin{aligned}
\max \quad & \frac{1}{3}\sum_{r=1}^{3}\theta_r + \varepsilon\sum_{i=1}^{2}s_i^- \\
\text{subject to:} \quad & \sum_{j=1}^{15}x_{1j}\lambda_j - s_1^- = x_{10}, \quad \sum_{j=1}^{15}x_{2j}\lambda_j - s_2^- = x_{20} \\
& w_1\sum_{\mathrm{j}=1}^{15}y_{1j}\lambda_j + w_2\sum_{j=1}^{15}y_{2j}\lambda_j + \ldots. + w_5\sum_{j=1}^{15}y_{5j}\lambda_j = \theta_1(w_1y_{10} + w_2y_{20} + \ldots + w_5y_{50}) \\
& \sum_{j=1}^{15}y_{6j}\lambda_j = \theta_2\ y_{60}, \qquad\qquad (5) \\
& \sum_{j=1}^{15}y_{7j}\lambda_j + \sum_{j=1}^{15}y_{8j}\lambda_j = \theta_3\ (y_{70} + y_{80}) \\
& \theta_r \geq 1,\ r = 1,2,3, \sum_{l=1}^{5}w_l = 1,\ 0.8 \leq \frac{w_l}{\overline{w}_l} \leq 1.2, \\
& \sum_{j=1}^{15}\lambda_j = 1, s_i^-, \lambda_j \geq 0,\ i = 1,2,\ j = 1,..,15,
\end{aligned}
$$

where $\overline{w}_l$ are the weights from the result of AHP in Table 1. Again, if the DMs prefer to compare the overall performance based on average level, then we have:

$$
\begin{aligned}
\max \quad & \frac{1}{3}\sum_{r=1}^{3}\theta_r + \varepsilon\sum_{i=1}^{2}s_i^- \\
\text{subject to:} \quad & \sum_{j=1}^{15}x_{1j}\lambda_j - s_1^- = x_{10}, \quad \sum_{j=1}^{15}x_{2j}\lambda_j - s_2^- = x_{20} \\
& w_1\sum_{\mathrm{j}=1}^{15}y_{1j}\lambda_j + w_2\sum_{j=1}^{15}y_{2j}\lambda_j + \ldots. + w_5\sum_{j=1}^{15}y_{5j}\lambda_j = \theta_1(w_1y_{10} + w_2y_{20} + \ldots + w_5y_{50}) \\
& \sum_{j=1}^{15}y_{6j}\lambda_j = \theta_2\ y_{60} \qquad\qquad (6) \\
& \sum_{j=1}^{15}y_{7j}\lambda_j + \sum_{j=1}^{15}y_{8j}\lambda_j = \theta_3\ (y_{70} + y_{80}) \\
& \sum_{l}^{5}w_l = 1, 0.8 \leq \frac{w_l}{\overline{w}_l} \leq 1.2, \theta_1 + \theta_2 + \theta_3 \geq 3, \\
& \sum_{j=1}^{15}\lambda_j = 1, s_i^-, \lambda_j, \theta_r \geq 0, i = 1,2, r = 1,2,3, j = 1,\ldots,15
\end{aligned}
$$

The results are presented in Model 6 column of Table 5. Comparing the results between Model 4 and 5, all the efficient DMUs are the same, while some inefficient DMUs are ranked in different order. The difference comes from different performance measurement. Taking Unit 4 as an example, the external research funding obtained of Unit 4 is far behind that of Unit 13, leading to a lower efficiency score when Russell measurement (Model 6) is used.

There are 11 DMUs classified as efficient by the BCC model, while the number of efficient DMUs drops down to 7 when multi-level DEA models are used. **It is clear that using the hierarchical structure can much increase the discrimination of DEA.**

With the average preference, only Unit 3 and 13 are still efficient, followed by Unit 2. Comparing with Rank M6 by Model 6 in Table 5 and Rank M 3 based on Model 3 in Table 3, the top 3 are same, while these three units were also ranked as the top three according to research sustainability evaluation in CES 2002. However, since only a part of research inputs was included in CES 2002, it is not much meaningful to compare the two results.

It seems possible to introduce the hierarchical structures in the DEA models of multiplier type by considering the following virtual sum:

$$\max \quad \frac{\left(u_1 \sum_{r=1}^{5} w_r \, y_{rj} + u_2 y_{6j} + u_3 \sum_{r=7}^{8} y_{rj} \right)}{\sum_{i=1}^{2} v_i x_{ij}}$$

where the multiplier u_s, v_i are the weights in the standard DEA multiplier models, but weights w_r reflect the hierarchical structures and satisfy the conditions used in the above models. Thus we have the three category groups, and inside these groups weighted compensations are allowed among the outputs. Again such a DEA model will be a more complicated nonlinear model. However since their dual models are not clear, it is not easy to clearly interpret their managerial meanings, although in some cases, these models could be re-written in the forms of multiplier DEA models with weight restrictions of ARI type. Also it does not seem to be straightforward to introduce Russell measurement and average preference in such models. However we think this direction is interesting and deserves attention.

Conclusions

In this paper, we carry out a pilot study on performance evaluation of 15 institutes for basic research in CAS using DEA considering the hierarchical structures of the indicators. The DMs often wish to use many indicators for relatively comprehensive evaluations in practical applications. Instead of removing some indicators according to statistical data correlation, in many cases it makes sense to group these indicators hierarchically, while different preferences may be preferred in different sub-systems to reflect their relative importance. The standard DEA models are not able to reflect such hierarchical structures, as they assume each indicator as equally important and non-substitutable. Meanwhile, it is also well known that discrimination of the standard DEA models will be diminishing if too many inputs and outputs are used. In order to cope with this problem, we look at multi-level DEA approaches, where indicators are constructed hierarchically. We also derive DEA models that can reflect hierarchical structures directly. We present a practical application of multi-level DEA models. It has been found that using the hierarchical structure can much increase the discrimination of DEA. In conclusion, we think it is feasible to apply DEA in future efficiency evaluation of CAS, and it is necessary to apply the DEA models that can take the value judgements of DMs into account.

Acknowledgement: The authors wish to express their sincere thanks to Bureau of Basic Research, Chinese Academy of Sciences for their assistances in writing this paper.

References:

1. Adler, N. and Golany, B. (2001) Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to Western Europe. *Journal of Operational Research*, 132(2), 260-273.
2. Cai, Y.Z. and Wu, W.J. (2001) Synthetic financial evaluation by a method of combining DEA with AHP. *International Transactions in Operational Research*, 8, 603-609.
3. Dyson, R.G., Allen, R., Camanho, A.S., Podinovski, V.V., Sarrico, C.S. and Shale, E.A. (2001) Pitfalls and protocols in DEA. *Europe Journal of Operational Research*, 132(2), 245-259.
4. Farzipoor Saen, R., Memariani, A. and Hosseinzadeh Lotfi, F. (2005) The effect of correlation coefficient among multiple input vectors on the efficiency mean in data envelopment analysis. *Applied Mathematics and Computation*, 162(2), 503-521.
5. Glänzel, W., Schubert, A. and Braun, T. (2002) A relational charting approach to the world of basic research in twelve science fields at the end of the second millennium. *Scientometrics,* 55(3), 335-348
6. Gracia, C.E. and Sanz-Menéndez, L. (2005) Competition for funding as an indicator of research competitiveness. *Scientometrics*, 64(3), 271-300.
7. Jenkins, L. and Anderson, M. (2003) A multivariate statistical approach to reducing the number of variables in data envelopment analysis. *European Journal of Operational Research,* 147, 51-61
8. Kao, C., Chang, P. and Hwang, S.N. (1993) Data envelopment analysis in measuring efficiency of forest management. *Journal of Environmental Management*, 38(1), 73-83.
9. Kaukonen, E. (1997) Evaluation of science research. In *Finland In The evaluation of scientific research: selected experiences*. OEVD/ED(97)194, Paris, OECD, 12-26.
10. King, D.A. (2004) The scientific impact of nations – what different countries get for their research spending. *Nature*, 430, 311-316.
11. Korhonen, P., Tainio, R. and Wallenius, J. (2001) Value efficiency analysis of academic research. *European Journal of Operational Research*, 130(1), 121-132.
12. Li, X.X. (2005) Practice and thoughts on performance evaluation in China’s state-run institutes of scientific research. *Bulletin of Chinese Academy of Sciences*, 20, 395-398.
13. Liu, F.H.F. and Hai, H.L. (2005) The voting analytic hierarchy process method for selecting supplier. *Journal of Production Economics*, 97(3), 308-317.

14. Liu, W.B., Meng, W., Zhang, T.Q. (2006) Chapter 21: Incorporating value judgements in DEA, 3rd edition of "Productivity Analysis in the Service Sector using Data Envelopment Analysis", 217-242.
15. Meng Wei. Studies of Data Envelopment Analysis and Its Applications in Basic Research Evaluation. University of Kent, PhD thesis, 2006.
16. Meng, W., Hu, Z.H. and Liu, W.B. (2006) Efficiency evaluation of basic research in China. *Scientometrics .Vol.69 85-101,2006.*
17. Meng, W., Zhang, D.Q., Liu, P.H., Peng, Z.L. and Zhang, W.Y. (2005) Analysis of a questionnaire about the CAS basic research intellectual property rights. *Bulletin of Chinese Academy of Sciences*, 20(4), 312-320.
18. Ramanathan, R. (2006) Data envelopment analysis for weight derivation and aggregation in the analytic hierarchy process. *Computers & Operations Research*, 33(5), 1289-1307
19. Shang, J. and Sueyposhi, T. (1995) A unified framework for the selection of a flexible manufacturing system. European Journal of Operational Research, 85(2), 297-315.
20. Sinuany-Sterm, A., Mehrez, A. and Barboy, A. (1994) Academic departments efficiency via DEA. *Computers and Operations Research*, 21(5), 543-556.
21. Takamura, Y. and Tone, K. (2003) A comparative site evaluation study for relocating Japanse government agencies out of Tokyo. *Socio-Economic Planning Science*, 37, 85-102.
22. Thanassoulis, E., Portela, M.C.A.S. and Allen, R. (2004) Incoporating value judgment in DEA. In W.W. Cooper, L.M. Seiford and J. Zhu (eds) *Handbook on Data Envelopment Analysis*. Kluwer Academic Publishers, Boston, 99-138.
23. Yang, T. and Kuo, C. (2003) A hierarchical AHP/DEA methodology for the facilities layout design problem. *European Journal of Operational Research*, 147(1), 128-136.
24. Zhang, Y. and Bartels, R. (1998) The effect of sample size on the mean efficiency in DEA with an application to electricity distribution in Australia, Sweden and New Zealand. J*ournal of Productive Analysis*, 9, 187-204.

University of Kent