Available online at www.sciencedirect.com
ScienceDirect
European Journal of Operational Research 180 (2007) 692–699
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/ejor

Decision Support

Classifying inputs and outputs in data envelopment analysis

Wade D. Cook [a,*], Joe Zhu [b]

[a] Department of Management Science, Schulich School of Business, York University, Toronto, Ont., Canada M3J 1P3
[b] Department of Management, Worcester Polytechnic Institute, Worcester, MA 01609, USA

Received 1 December 2003; accepted 29 March 2006
Available online 30 June 2006

Abstract

In conventional data envelopment analysis it is assumed that the input versus output status of each of the chosen performance measures is known. In some situations, however, certain performance measures can play either input or output roles. We refer to these performance measures as flexible measures. This paper presents a modification of the standard constant returns to scale DEA model to accommodate such flexible measures. Both an individual DMU model and an aggregate model are suggested as methodologies for deriving the most appropriate designations for flexible measures. We illustrate the application of these models in two practical problem settings.

Keywords: DEA; Variable status; Flexible inputs; Flexible outputs; Efficiency

1. Introduction

Data envelopment analysis (DEA), developed by Charnes et al. (1978), provides a nonparametric methodology for evaluating the efficiency of each of a set of comparable decision making units (DMUs), relative to one another. In the original model of Charnes et. al., efficiency is represented by the ratio of weighted outputs to weighted inputs. An important feature of DEA is its capability to provide efficiency scores, while taking account of both multiple inputs and multiple outputs. This methodology has been applied in a wide range of applications over the past 25 years, in settings that include hospitals, banks, maintenance crews, etc., see Cooper et al. (2004).

In the conventional application of DEA, it is assumed that one can, given a collection of available measures, clearly specify which will constitute inputs and which outputs. For example, in a conventional study of efficiency of bank branch operations, such as discussed in Cook et al. (2000, 2001), outputs used are the standard counter transactions such as deposits and withdrawals, and inputs are resources such as various staff types. Suppose, however, that one wishes to evaluate each branch's efficiency to attract investments. In this case, a factor such as the number of "high value" customers, could serve as either an input or an output. From one perspective, such a measure may play the role of proxy for future investment, hence can reasonably be classified as an output. On the other hand, it can legitimately be considered as an environmental input that aids the

* Corresponding author. Tel.: +1 416 7365074; fax: +1 416 7365687.
E-mail addresses: wcook@schulich.yorku.ca (W.D. Cook), jzhu@wpi.edu (J. Zhu).

0377-2217/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2006.03.048

branch in generating its existing investment portfolio. A case can be made, as well, for measures such as deposits. In some bank previous studies, such a measure is regarded as an output, in that it is a source of revenue for the branch. At the same time, arguments have been made that staff time expended in processing customers who are making deposits or opening deposit accounts, could be used to better advantage to sell more profitable products to the customer. Similar arguments can be made regarding the evaluation of research productivity by universities, such as described in Beasley (1990, 1995). There, "research income" is treated as both an output and input. A related problem setting is that in which research – granting agencies (e.g., NSERC in Canada and NSF in the USA) wish to allocate funds to those researchers and universities such as to have the greatest impact. In this environment, graduate students can play the role of either an input (a resource available to the faculty member, effecting his/her productivity), or as an output (trained personnel, hence a benefit resulting from research funding). Nurse trainees and medical interns have a similar interpretation in the evaluation of hospital efficiency. In a very different environment, Cook et al. (1990), in evaluating highway maintenance crew efficiency, use the measure "average pavement rating" as an input that (negatively) influences the outputs. At the same time, one might make the argument that this measure is, as well, an output that is clearly influenced by the level of annual maintenance expenditure.

One might argue that choosing the appropriate status for a variable, when ambiguity prevails, has mainly to do with providing the fairest treatment possible for the individual DMU. At the same time, if one views performance measurement from the perspective of the overseeing organization as the 'decision maker' or manager of the DMUs, such ambiguity may be even more of an issue. In the case of the research granting agency, the decision as to whether graduate students should be designated as an input or output can have very real consequences on the funds received by the individual applicants. Hence, it is in the interest of the agency to use the fairest and hopefully least controversial method possible to evaluate efficiency.

In many problem situations such as those described, the input versus output status of certain measures can be deemed as flexible. In recent research by Cook and Bala (2003), the problem of deciding the appropriate status of such measures was examined when additional information is present. Specifically, they investigate the situation where bank branch consultants provide additional "classification" data specifying which branches, in their assessment, qualify as good versus poor branches. The idea there was to assign a status to each flexible variable such as to provide efficiency scores that are in best agreement with expert opinion. In the current paper we provide a methodology for deciding the status of flexible variables in those settings where such flexibility is present.

In Section 2 we develop an augmentation to the conventional DEA structure, to facilitate the derivation of the input versus output status of some variables, when the flexibility of those variables is an issue. Section 3 applies the developed models in two settings. First we apply the ideas to the data of Beasley (1990, 1995) to determine the appropriate classification of "research income". In Beasley's case, he allows this variable to occupy the roles of both input and output, simultaneously. Next, we apply the augmented models to a multiple factor case involving bank branch data as per Cook et al. (2000, 2001, 2004). Conclusions and further research are discussed in Section 4.

2. Identifying the Input Output Status of Flexible Variables

Suppose we wish to evaluate the efficiencies of n decision making units (DMUs). Each DMU_j, $j=1,\ldots,n$ produces s different outputs y_{rj} ($r=1,2,\ldots,s$), using m different inputs x_{ij} ($i=1,2,\ldots,m$). The CCR ratio model of Charnes et al. (1978), is given by:

$$\begin{aligned} \max \quad & \frac{\sum_{r=1}^{s}\mu_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}} \\ \text{s.t.} \quad & \frac{\sum_{r=1}^{s}\mu_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leqslant 1, \quad j=1,2,\ldots,n \\ & \mu_r, v_i \geqslant 0, \quad \forall r,i \end{aligned} \tag{1}$$

Suppose also that there exist L "flexible measures", whose input/output status we wish to determine. We denote the values assumed by these measures as w_{lj} for DMU_j ($l=1,..,L$). For each measure l, we introduce the binary variables $d_l \in \{0,1\}$, where $d_l=1$ designates that factor l is an output, and $d_l=0$ designates it as an input.

Let γ_l be the weight for each measure l. We establish the following mathematical programming model:

$$\max \quad \frac{\sum_{r=1}^{s}\mu_r y_{ro} + \sum_{l=1}^{L} d_l\gamma_l w_{lo}}{\sum_{i=1}^{m} v_i x_{io} + \sum_{l=1}^{L}(1-d_l)\gamma_l w_{lo}}$$
$$\text{s.t.} \quad \frac{\sum_{r=1}^{s}\mu_r y_{rj} + \sum_{l=1}^{L} d_l\gamma_l w_{lj}}{\sum_{i=1}^{m} v_i x_{ij} + \sum_{l=1}^{L}(1-d_l)\gamma_l w_{lj}} \leqslant 1, \quad j=1,2,\ldots,n \tag{2}$$
$$d_l \in \{0,1\},\ \forall l \quad \mu_r, v_i, \gamma_l \geqslant 0, \quad \forall r,i,l$$

Note that since either $d_l = 0$ or $d_l = 1$, then measure l will end up either as an output or an input. Thus, problem (2) for DMU_o permits that DMU to choose which is best for each measure l, whether to designate it as an output or input.

For any fixed value d_l (whether 0 or 1), model (2) becomes the following programming problem:

$$\begin{aligned}
\max \quad & \sum_{r=1}^{s}\mu_r y_{ro} + \sum_{l=1}^{L} d_l\gamma_l w_{lo} \\
\text{s.t.} \quad & \sum_{r=1}^{s} u_r y_{rj} + \sum_{l=1}^{L} d_l\gamma_l w_{lj} - \sum_{i=1}^{s} v_i x_{ij} \\
& \quad - \sum_{l=1}^{L}(1-d_l)\gamma_l w_{lj} \leqslant 0, \quad j=1,\ldots,n; \\
& \sum_{i=1}^{m} v_i x_{io} + \sum_{l=1}^{L}(1-d_l)\gamma_l w_{lo} = 1 \\
& d_l \in \{0,1\}, \quad \mu_r, v_i, \quad \gamma_l \geqslant 0, \quad \forall r,i,l
\end{aligned} \tag{3}$$

Model (3) is clearly nonlinear. It can, however, be linearized by way of the change of variables $\delta_l = d_l\gamma_l$, and, for each l, imposing the constraints

$$\begin{aligned}
& 0 \leqslant \delta_l \leqslant M d_l \\
& \delta_l \leqslant \gamma_l \leqslant \delta_l + M(1-d_l)
\end{aligned} \tag{4}$$

where M is a large positive number. These linear integer restrictions capture the nonlinear expression $d_l\gamma_l = \delta_l$, without actually having to directly specify it in the optimization model. Note that if $d_l = 1$ then $\gamma_l = \delta_l$, and if $d_l = 0$, then $\delta_l = 0$. We, therefore, have the following mixed integer linear program:

$$\begin{aligned}
\max \quad & \sum_{r=1}^{s}\mu_r y_{ro} + \sum_{l=1}^{L} \delta_l w_{lo} \\
\text{s.t.} \quad & \sum_{r=1}^{s}\mu_r y_{rj} + 2\sum_{l=1}^{L} \delta_l w_{lj} - \sum_{i=1}^{s} v_i x_{ij} \\
& \quad - \sum_{l=1}^{L}\gamma_l w_{lj} \leqslant 0, \quad j=1,\ldots,n; \\
& \sum_{i=1}^{m} v_i x_{io} + \sum_{l=1}^{L}\gamma_l w_{lo} - \sum_{l=1}^{L}\delta_l w_{lo} = 1 \\
& 0 \leqslant \delta_l \leqslant M d_l \\
& \delta_l \leqslant \gamma_l \leqslant \delta_l + M(1-d_l) \\
& d_l \in \{0,1\}, \quad \delta_l, \gamma_l \geqslant 0, \quad \forall l; \quad \mu_r, v_i \geqslant 0, \ \forall r,i
\end{aligned} \tag{5}$$

There appear to be at least two possible approaches for deciding the status of the flexible variables in a DEA setting. The first and most obvious approach is to examine the issue from the point of view of the individual DMU. Specifically, we would solve model (5) for each DMU, and obtain a set of optimal d_l^* specific to that DMU. One criterion for deciding the overall input versus output status of any flexible measure would then be to base it on the majority choice among the DMUs. The use of a simple majority decision rule is widely applied in many settings, and would seem to be the least controversial way to make the choice here. An alternative approach would be to look at the situation from the perspective of the manager of the collection of DMUs. Specifically, use the designation (input or output) for each flexible variable that renders the *aggregate efficiency* of the collection of DMUs as large as possible. Such a model would be helpful if ties are encountered using model (5) on an individual DMU basis. In the development below we adopt this aggregate approach.

We, therefore, assume that the optimal input/output designation for flexible variables, will be the one created by the model which optimizes the aggregate or average ratio of outputs to inputs, namely

$$\max \quad \frac{\sum_{r=1}^{s}\mu_r\left(\sum_{j=1}^{n} y_{rj}\right) + \sum_{l=1}^{L} d_l\gamma_l\sum_{j=1}^{n} w_{lj}}{\sum_{i=1}^{m} v_i\left(\sum_{j=1}^{n} x_{ij}\right) + \sum_{l=1}^{L}(1-d_l)\gamma_l\sum_{j=1}^{n} w_{lj}}$$
$$\text{s.t.} \quad \frac{\sum_{r=1}^{s}\mu_r y_{rj} + \sum_{l=1}^{L} d_l\gamma_l w_{lj}}{\sum_{i=1}^{m} v_i x_{ij} + \sum_{l=1}^{L}(1-d_l)\gamma_l w_{lj}} \leqslant 1, \quad j=1,2,\ldots,n$$
$$d_l \in \{0,1\}, \quad \mu_r, v_i, \gamma_l \geqslant 0, \quad \forall r,i,l \tag{6}$$

For notational convenience, let $\tilde{y}_r = \sum_{j=1}^{n} y_{rj}$, $\tilde{x}_i = \sum_{j=1}^{n} x_{ij}$, and $\tilde{w}_l = \sum_{j=1}^{n} w_{lj}$. Using (4), our aggregate problem is then equivalent to the following integer linear programming problem:

$$\begin{aligned}
\max \quad & \sum_{r=1}^{s}\mu_r \tilde{y}_r + \sum_{l=1}^{L} \delta_l \tilde{w}_l \\
\text{s.t.} \quad & \sum_{r=1}^{s}\mu_r y_{rj} + 2\sum_{l=1}^{L} \delta_l w_{lj} - \sum_{i=1}^{s} v_i x_{ij} \\
& \quad - \sum_{l=1}^{L}\gamma_l w_{lj} \leqslant 0, \quad j=1,\ldots,n \\
& \sum_{i=1}^{m} v_i \tilde{x}_i + \sum_{l=1}^{L}\gamma_l \tilde{w}_l - \sum_{l=1}^{L}\delta_l \tilde{w}_l = 1 \\
& 0 \leqslant \delta_l \leqslant M d_l \\
& \delta_l \leqslant \gamma_l \leqslant \delta_l + M(1-d_l) \\
& d_l \in \{0,1\}, \quad \mu_r, v_i, \gamma_l \geqslant 0, \quad \forall r,i,l
\end{aligned} \tag{7}$$

In the following section we apply these developments to aid in determining the input output status of flexible variables in two real world problem settings.

3. Application

In this section, we apply our models to the data sets used in Beasley (1990) (see Table 1) and Cook

Table 1
University data (Beasley, 1990)

DMU	General expenditure	Equipment expenditure	UG students	PG teaching	PG research	Research income
University1	528	64	145	0	26	254
University2	2605	301	381	16	54	1485
University3	304	23	44	3	3	45
University4	1620	485	287	0	48	940
University5	490	90	91	8	22	106
University6	2675	767	352	4	166	2967
University7	422	0	70	12	19	298
University8	986	126	203	0	32	776
University9	523	32	60	0	17	39
University10	585	87	80	17	27	353
University11	931	161	191	0	20	293
University12	1060	91	139	0	37	781
University13	500	109	104	0	19	215
University14	714	77	132	0	24	269
University15	923	121	135	10	31	392
University16	1267	128	169	0	31	546
University17	891	116	125	0	24	925
University18	1395	571	176	14	27	764
University19	990	83	28	36	57	615
University20	3512	267	511	23	153	3182
University21	1451	226	198	0	53	791
University22	1018	81	161	5	29	741
University23	1115	450	148	4	32	347
University24	2055	112	207	1	47	2945
University25	440	74	115	0	9	453
University26	3897	841	353	28	65	2331
University27	836	81	129	0	37	695
University28	1007	50	174	7	23	98
University29	1188	170	253	0	38	879
University30	4630	628	544	0	217	4838
University31	977	77	94	26	26	490
University32	829	61	128	17	25	291
University33	898	39	190	1	18	327
University34	901	131	168	9	50	956
University35	924	119	119	37	48	512
University36	1251	62	193	13	43	563
University37	1011	235	217	0	36	714
University38	732	94	151	3	23	297
University39	444	46	49	2	19	277
University40	308	28	57	0	7	154
University41	483	40	117	0	23	531
University42	515	68	79	7	23	305
University43	593	82	101	1	9	85
University44	570	26	71	20	11	130
University45	1317	123	293	1	39	1043
University46	2013	149	403	2	51	1523
University47	992	89	161	1	30	743
University48	1038	82	151	13	47	513
University49	206	1	16	0	6	72
University50	1193	95	240	0	32	485

and Zhu (2005). See also Cook et al. (2000). Let us first consider a portion of the data set in Beasley, where we use two inputs, General Expenditure and Equipment Expenditure and three outputs, consisting of three types of students. The "flexible measure" here is the Research Income. We note that in Beasley's analysis, this latter variable was treated as both an input and output.

Table 2
Results from model (5)

DMU	Deposits	Efficiency
University1	1	1
University2	0	1
University3	0	0.837244
University4	1	0.685697
University5	0	1
University6	0	1
University7	1	1
University8	1	0.811941
University9	0	1
University10	1	0.906595
University11	0	0.890126
University12	1	0.709313
University13	0	0.803249
University14	0	0.767744
University15	0	0.704214
University16	0	0.54274
University17	1	0.819451
University18	1	0.627824
University19	1	1
University20	0	1
University21	0	0.699625
University22	1	0.716738
University23	0	0.617112
University24	0	1
University25	1	1
University26	0	1
University27	1	0.855471
University28	0	1
University29	1	0.824968
University30	0	1
University31	1	0.775853
University32	0	0.896402
University33	1	1
University34	0	1
University35	1	1
University36	0	0.8369
University37	1	0.830789
University38	0	0.833414
University39	0	0.791219
University40	1	0.741404
University41	1	1
University42	0	0.847172
University43	0	0.920638
University44	0	1
University45	0	1
University46	0	1
University47	1	0.688445
University48	0	0.938878
University49	0	1
University50	0	0.841683

Table 2 reports the results from model (5), where the second column shows the optimal d and the third column, the optimal value to model (5). In this case, 20 out of the 50 universities treat the research income measure as an output, i.e., the majority of 30 treat it as an input.

When we apply our aggregate model (7), the optimal $d = 1$, with the optimal objective function value being 0.69329. This indicates that from an aggregate efficiency perspective, research income is treated as an output. Part of the explanation for the different results between the two approaches may be that for the 30 cases where the input was the preferred status using model (7), that preference may not have been as strong as was the preference of output versus input in the other 20 cases. Hence, the majority concept may be flawed by failing to take account of the difference in efficiency for each DMU, when the flexible measure is used as an output versus an input. At the same time, it is recognized that the aggregate approach might be criticized for being overly sensitive to extreme DMUs, or possibly to the 'larger' DMUs. Such arguments are common in the MCDM literature, where a consensus ranking among responses from a collection of voters is to be derived. See, for example, Cook (2006). The aggregate approach given here is in the spirit of many consensus ranking methods.

We next applied our models to a bank data set used in Cook and Zhu (2005). Because the data set covers 100 branches, we do not provide it here. Table 3 displays inputs and outputs for a sample of branches. It is noted that there are gaps in the numbering of the branches due to the fact that the sample here was drawn from a larger population. In the first analysis, "deposits" is the only flexible measure. Table 4 provides the results from this analysis. We note that 89 out of 100 branches treat deposits as an output. When we apply the data set to model (7), the optimal $d = 1$, with the optimal value to model (7) being 0.858, indicating that the "deposits" is again treated as an output.

Finally, Table 5 displays the results arising from model (5) when three flexible variables ("deposits", "open account", and "withdrawals") are considered. As can be seen, various combinations occur. For example, branch #1 treats "deposits" as an out-

Table 3
Sample data for bank branch analysis

Branch	Inputs			Outputs								
	FTESales	FTESer	FTEOth	Deposits	OpenAcct	WD	UPD	TRF	VISA	RRSP	Let Cr.	Loans
1	4.95	7.7	10.605	87,649	841	3981	2636	88	551	990	820	431
2	18.15	16.5	35.515	220,726	2729	12019	6653	213	2973	4210	1370	1132
3	4.125	4.4	3.63	38,912	431	2437	992	4	422	880	440	304
4	3.3	4.4	3.96	41,258	724	2342	1421	35	355	940	80	412
5	1.65	2.2	2.64	33,658	260	942	854	117	100	140	70	157
6	4.125	4.4	6.5956	58,942	533	1832	1281	213	226	820	350	373
7	3.3	4.4	8.0462	56,061	752	2320	1044	202	710	750	480	214
8	4.125	4.4	4.4061	46,968	872	3005	1441	429	578	850	160	402
9	9.075	9.9	15.186	130,035	2001	3878	1880	1094	780	2630	1370	675
10	12.375	18.7	15.327	155,202	4731	6502	6349	1942	834	3010	2290	915

Table 4
Model (5) results for a single flexible variable – deposits

Bank	Deposits	Efficiency	Bank	Deposits	Efficiency	Bank	Deposits	Efficiency
1	1	1.000	56	1	0.815	105	1	1.000
2	1	1.000	57	1	1.000	107	1	1.000
3	1	0.999	58	1	0.998	109	1	0.946
4	1	1.000	59	1	0.975	111	1	1.000
5	1	1.000	60	1	1.000	112	1	1.000
6	1	1.000	61	1	0.947	113	0	1.000
7	1	1.000	62	1	0.874	115	1	1.000
8	1	1.000	63	1	0.983	116	1	1.000
9	1	1.000	64	1	1.000	117	1	1.000
10	1	1.000	66	0	0.936	120	1	1.000
14	1	1.000	67	1	0.872	122	1	1.000
16	1	1.000	68	0	1.000	123	1	1.000
18	1	1.000	70	0	1.000	125	0	0.904
20	1	1.000	71	0	1.000	127	1	1.000
21	1	1.000	72	1	1.000	128	1	0.947
23	1	1.000	74	0	1.000	131	1	0.858
24	1	0.940	75	1	1.000	132	1	1.000
27	1	1.000	76	1	1.000	133	1	1.000
28	1	1.000	77	1	0.970	135	1	1.000
29	1	1.000	78	1	1.000	137	1	0.966
30	1	1.000	79	0	0.841			
31	1	0.847	80	1	1.000			
33	1	1.000	81	1	1.000			
34	1	1.000	82	1	0.838			
36	1	1.000	83	1	1.000			
39	1	1.000	86	1	0.966			
40	1	0.942	87	1	1.000			
42	1	0.898	88	1	1.000			
43	1	0.981	89	1	1.000			
44	0	1.000	90	1	1.000			
45	1	0.989	91	1	1.000			
46	1	0.933	92	1	1.000			
47	1	0.833	94	1	1.000			
48	0	0.755	96	1	1.000			
49	1	1.000	97	1	0.980			
50	1	1.000	98	0	0.942			
51	1	1.000	99	1	1.000			
52	1	1.000	100	1	1.00			
53	1	1.000	102	1	0.97			
55	1	1.000	103	1	1.00			

Table 5
Results from model (5) for three flexible bank branch variables

Branch	d1 (Deposits)	d2 (Open)	d3 (WD)	New Efficiency	Old Efficiency	Branch	d1 (Deposits)	d2 (Open)	d3 (WD)	New Efficiency	Old Efficiency
1	1	0	0	1.000	1.000	67	0	1	0	0.919	0.872
2	0	0	0	1.000	1.000	68	0	1	0	0.990	0.923
3	1	0	0	1.000	0.999	70	0	1	1	1.000	0.999
4	0	0	0	1.000	1.000	71	0	1	0	1.000	0.985
5	0	0	0	0.928	1.000	72	0	1	0	0.908	1.000
6	0	0	0	1.000	1.000	74	0	1	0	1.000	0.966
7	1	0	0	1.000	1.000	75	1	0	0	1.000	1.000
8	0	1	0	1.000	1.000	76	1	0	0	1.000	1.000
9	0	1	0	1.000	1.000	77	0	1	1	0.937	0.970
10	0	1	0	1.000	1.000	78	0	1	0	0.982	1.000
14	0	0	1	1.000	1.000	79	0	1	0	0.843	0.801
16	1	0	0	1.000	1.000	80	0	0	1	1.000	1.000
18	1	1	0	1.000	1.000	81	0	1	0	1.000	1.000
20	1	0	0	1.000	1.000	82	1	1	0	0.855	0.838
21	0	0	0	1.000	1.000	83	0	1	0	1.000	1.000
23	0	1	0	1.000	1.000	86	0	0	0	0.946	0.966
24	0	1	0	0.910	0.940	87	0	1	0	1.000	1.000
27	1	0	0	1.000	1.000	88	0	0	1	1.000	1.000
28	0	1	0	1.000	1.000	89	0	1	0	1.000	1.000
29	1	0	0	1.000	1.000	90	1	0	0	1.000	1.000
30	1	0	0	1.000	1.000	91	0	1	0	1.000	1.000
31	1	0	0	1.000	0.847	92	0	0	1	1.000	1.000
33	0	0	1	1.000	1.000	94	0	1	0	1.000	1.000
34	0	1	0	1.000	1.000	96	0	0	1	0.996	1.000
36	1	0	0	1.000	1.000	97	0	1	0	0.975	0.980
39	0	1	0	1.000	1.000	98	0	1	0	1.000	0.935
40	1	0	0	1.000	0.942	99	0	1	1	0.960	1.000
42	1	1	0	0.913	0.898	100	1	0	0	1.000	1.000
43	0	1	0	1.000	0.981	102	0	1	0	1.000	0.978
44	1	0	0	1.000	0.948	103	0	0	1	1.000	1.000
45	0	1	0	1.000	0.989	105	1	0	0	1.000	1.000
46	1	1	0	0.933	0.933	107	1	0	0	1.000	1.000
47	1	1	0	0.845	0.833	109	1	0	0	0.947	0.946
48	0	0	0	0.917	0.703	111	1	0	0	1.000	1.000
49	0	1	0	1.000	1.000	112	1	0	0	1.000	1.000
50	1	0	0	1.000	1.000	113	0	0	1	1.000	0.988
51	0	1	0	1.000	1.000	115	1	0	0	1.000	1.000
52	0	1	0	0.917	1.000	116	0	0	1	1.000	1.000
53	1	0	0	1.000	1.000	117	0	0	1	1.000	1.000
55	0	0	1	1.000	1.000	120	0	1	0	1.000	1.000
56	1	0	0	1.000	0.815	122	1	0	0	1.000	1.000
57	0	1	1	1.000	1.000	123	0	1	0	1.000	1.000
58	0	1	0	0.957	0.998	125	0	1	0	0.953	0.829
59	1	1	0	0.973	0.975	127	0	1	0	0.939	1.000
60	0	1	0	1.000	1.000	128	0	1	0	1.000	0.947
61	0	0	0	0.932	0.947	131	1	1	0	0.890	0.858
62	1	0	1	0.891	0.874	132	0	0	0	1.000	1.000
63	0	1	0	0.917	0.983	133	1	0	0	1.000	1.000
64	1	0	0	1.000	1.000	135	0	1	0	1.000	1.000
66	0	0	0	1.000	0.878	137	0	1	0	0.909	0.966

put, but the other two factors as inputs. Branch #18, however, treats the first two variables as outputs and only the last one as an input. When we apply model (7) in this situation the optimal aggregate efficiency score of 0.8756 results from the first two measures, ‘‘deposits’’, and ‘open account’’ are designated as outputs, while the latter, ‘‘withdrawals’’ is designated an input.

4. Conclusions

Conventional DEA analyses require that each variable or measure be assigned an explicit designation specifying whether it is an input or output. In various settings, however, it remains that there are variables whose status is flexible. In cases where a resource can as well represent a tangible product of the organization (nurse trainees, medical interns, graduate students, research funding, etc.), this flexibility is present.

In the current paper we develop a modification to the standard constant returns to scale model that permits the inclusion of such flexible measures in the analysis. We believe this to be a significant contribution to an important and very much under-researched topic. The model assigns an optimal designation, whether input or output to each such variable. Two models are given for accomplishing this, namely an individual DMU model, and one that optimizes the aggregate efficiency of the collection of DMUs. We apply these models in two applied settings.

We emphasize again that we are assuming that the only candidates for flexible variables are those whose status can clearly be either that of an input or output. Variables such as cost, that are clearly inputs, would never be subjected to the analysis herein. As well, we argue that choosing the status of a flexible variable by adopting a 'majority rule' argument is consistent with the many models discussed in the MCDM literature as per Cook (2006).

We note that both the individual and aggregate methodologies (models (5) and (7)), may decrease the discriminatory power of the resulting DEA model with respect to the number of efficient DMUs generated. From an optimization perspective, it is natural that this should occur. However, sufficient discrimination among the efficiency scores in any DEA analysis is often an issue. In those cases (and for the setting herein) such discrimination is generally achieved by way of imposed restrictions on the variable multipliers. The assurance region (AR) methodology of Thompson et al. (1990) provides for one such means of applying multiplier constraints.

References

Beasley, J., 1990. Comparing university departments. Omega 8 (2), 171–183.

Beasley, J., 1995. Determining teaching and research efficiencies. Journal of the Operational Research Society 46, 441–452.

Charnes, A., Cooper, W., Rhodes, E., 1978. Measuring the efficiency of decision making units. European Journal of Operational Research 2 (6), 428–444.

Cook, W.D., 2006. Distance-based and ad hoc consensus models in ordinal preference ranking. European Journal of Operational Research 172 (2), 369–385.

Cook, W.D., Bala, K., 2003. Performance measurement with classification information: An enhanced additive DEA model. Omega 31 (6), 439–450.

Cook, W.D., Hababou, M., 2001. Sales performance measurement in bank branches. Omega 29, 299–307.

Cook, W.D., Zhu, J., 2005. Building performance standards into DEA structures. IIE Transactions 37 (3), 267–275.

Cook, W.D., Kazakov, A., Roll, Y., 1990. A DEA model for measuring the relative efficiency of highway maintenance patrols. Infor 28 (2), 113–124.

Cook, W.D., Hababou, M., Tuenter, H., 2000. Multicomponent efficiency measurement and shared inputs in data envelopment analysis: An application to sales and service performance in bank branches. Journal of Productivity Analysis 14 (3), 209–224.

Cook, W.D., Seiford, L., Zhu, J., 2004. Models for performance benchmarking: Measuring the effect of e-business activities on banking performance. Omega 32 (4), 313–322.

Cooper, W.W., Seiford, L.M., Zhu, J., 2004. Handbook of Data Envelopment Analysis. Kluwer Academic Publishers., Norwell, MA.

Thompson, R.G., Langemeier, L.N., Lee, C.T., Lee, E., Thrall, R.M., 1990. The role of multiplier bounds in efficiency analysis with application to Kansas farming. Journal of Econometrics 46, 93–108.