A New Framework for the solution of DEA models

Gautam Appa and H. Paul Williams

London School of Economics and Political Science

A New Framework for the solution of DEA models

Gautam Appa and H. Paul Williams

London School of Economics

This working paper is a continuation of our earlier working paper 'A formula for the solution of DEA models' LSEOR 02.49. The same illustrative example is used, but many more applications are included.

Abstract

We provide an alternative framework for solving Data Envelopment Analysis (DEA) models which, in comparison with the standard Linear Programming (LP) based approach that solves one LP for each Decision Making Unit (DMU), delivers much more information. By projecting out all the variables which are common to all LP runs, we obtain a formula into which we can substitute the inputs and outputs of each DMU in turn in order to obtain its efficiency number and all possible primal and dual optimal solutions. The method of projection, which we use, is Fourier-Motzkin (F-M) Elimination. This provides us with the finite number of extreme rays of the elimination cone. These rays give the dual multipliers which can be interpreted as weights which will apply to the inputs and outputs for particular DMUs. As the approach provides *all* the extreme rays of the cone, multiple sets of weights, when they exist, are explicitly provided. Several applications are presented. It is shown that the output from the F-M method improves on existing methods of i) establishing the returns to scale status of each DMU, ii) calculating cross-efficiencies and iii) dealing with weight flexibility. The method also demonstrates that the same weightings will apply to all DMUs having the same comparators. In addition it is possible to construct the *skeleton* of the efficient frontier of efficient DMUs. Finally, our experiments clearly indicate that the extra computational burden is not excessive for most practical problems.

Subject Classifications: *Data Envelopment Analysis, Returns to Scale in DEA, Cross-efficiency in DEA, Weight Restrictions in DEA, Fourier-Motzkin elimination.*

1. INTRODUCTION

DEA is an established method of comparing the performance of a number of similar units (eg shops, hospitals, schools, garages etc). Many case studies have been published (see eg Seiford (1995), Cooper et al (1999)).

The individual units are known as Decision Making Units (DMUs). Each DMU has a number of *inputs* used to produce several *outputs*. In the standard model it is assumed that there are *constant returns to scale* (eg doubling all inputs will result in the doubling of all outputs). This implies *infinite divisibility* (ie fractional amounts of the inputs can be used). In addition *additivity* is also assumed, enabling one to make up fictional 'virutal' DMUs by combining actual DMUs in any multiples. All these assumptions can be altered to produce non-standard models. A full description of DEA can be found in many texts (eg Cooper et al (1999), Thanassoulis (2001)). For the purposes of this paper we will use the standard model, known in the literature as the CCR model, first introduced in the pioneering work of Charnes et al (1978).

In section 2 we formulate the standard model as an LP and show how it is possible to *project out* all but one of the variables to give us weightings for the inputs and outputs, resulting in a *formula* which gives us the efficiency number as well as a complete set of optimal weights and comparators for each DMU. It also enables us to construct the *skeleton* of the efficiency frontier. This is illustrated, in section 3, by a numerical example. Section 4 outlines several applications. It is shown that the output from the F-M method:

    i)   provides the returns to scale status of all the DMUs;

    ii)   makes it unnecessary to use surrogate models (as done by Doyle and Green (1994) and Sexton (1986)) to obtain the optimal choice of weights for each DMU that maximises its own efficiency while minimising or maximising the average of the cross-efficiencies of its peers;

    iii)   provides a tool for selecting exemplary DMUs for the Polyhedral Cone-Ratio model of weight restrictions by furnishing the range of variations in efficiency levels for alternative selections.

Finally, in section 5, we attend to the extra computing time needed for our approach and look at its viability for solving normal and larger practical problems as well as suggest ideas on how this approach may be extended to non-standard models.

2. PROJECTION OF THE STANDARD MODEL

We consider a problem with *n* DMUs indexed by j in {1,2,…,n}. Each DMU has *p* inputs and *q* outputs. DMUj has inputs x_{1j}, x_{2j},…,x_{pj} and outputs y_{1j}, y_{2j},…,y_{qj}.

In the *Primal* model we introduce the following variables:

λ_j = Amount of DMUj used;

θ = The proportion of the input bundle of DMUk needed to produce its output bundle.

In order to evaluate DMUk we have the model:

Minimise θ

subject to: $- x_{11}\,\lambda_1 - x_{12}\,\lambda_2 - \ldots - x_{1n}\,\lambda_n + x_{1k}\,\theta \geq 0$
  $- x_{21}\,\lambda_1 - x_{22}\,\lambda_2 - \ldots - x_{2n}\,\lambda_n + x_{2k}\,\theta \geq 0$

  $.\qquad\quad .\qquad\quad .\qquad\quad .\qquad\quad .$

  $- x_{p1}\,\lambda_1 - x_{p2}\,\lambda_2 - \ldots - x_{pn\,\lambda n} + x_{pk}\,\theta \geq 0$
  $y_{11}\,\lambda_1 + y_{12}\,\lambda_2 + \ldots + y_{1n}\,\lambda_n \qquad \geq y_{1k}$
  $y_{21}\,\lambda_1 + y_{22}\,\lambda_2 + \ldots + y_{2n}\,\lambda_n \qquad \geq y_{2k}$

  $.\qquad\quad .\qquad\quad .\qquad\quad .\qquad\quad .$

  $y_{q1}\,\lambda_1 + y_{q2}\,\lambda_2 + \ldots + y_{qn}\,\lambda_n \qquad \geq y_{qk}$
  $\lambda_j \geq 0 \quad j = 1,2,\ldots,n$

Here x_{ij} is the amount of input i used by DMUj for i = 1, … , p and

y_{tj} is the amount of output t produced by DMUj for t = 1, … , q.

Its interpretation is as follows:

Choose a mixture of DMUs to produce at least the outputs of DMUk using the smallest possible multiple of the inputs of DMUk. If DMUk is *efficient*, its outputs will be best produced using all of its own inputs. In this case $\lambda_k = 1$, $\lambda_j = 0$ for all j ≠ k and θ =1 (θ is its efficiency number). On the other hand, if DMUk is *inefficient*, its outputs will be best produced by a mixture of *other* DMUs using a fraction θ of all its inputs. In this case θ, its efficiency number, is such that $0 < \theta < 1$.

If we *project* out variables λ_1, λ_2, … , λ_n using F- M elimination (see Williams (1986)) this results in the model taking the form:

$$\left(\sum_{i=1}^{p} u_{il}\ x_{ik} \right) \theta \geq \sum_{t=1}^{q} v_{tl}\ y_{tk} \quad \text{for } l = 1,2, \ldots L$$

The optimal value of θ is

$$\theta^* = \underset{l}{Max}\left[\left(\sum_{t=1}^{q} v_{tl}\ y_{tk} \right) \Big/ \left(\sum_{i=1}^{p} u_{il}\ x_{ik} \right) \right].$$

Let $\mathrm{arg}\max\limits_{l} \theta^* = m$ (1)

θ^* gives DMUk's efficiency number and u_{im}, v_{tm} are the weightings which it should use, for its inputs and outputs, in order to maximise its ratio of weighted outputs to inputs.

It is convenient to consider the inputs and outputs as a matrix.

$$A = \begin{bmatrix} -x_{11} & -x_{12} \, ... -x_{1n} \\ -x_{21} & -x_{22} \, ... -x_{2n} \\ ... & ... \\ -x_{p1} & -x_{p2} \, ... -x_{pn} \\ y_{11} & y_{12} \, \cdots \, y_{1n} \\ y_{21} & y_{22} \, \cdots \, y_{2n} \\ ... & ... \\ y_{q1} & y_{q2} \, \cdots \, y_{qn} \end{bmatrix}$$

In order to eliminate the variables $\lambda_1, \lambda_2, ..., \lambda_n$ we consider the matrix **B**, with non-negative rows, consisting of the (finite number of) extreme rays of the homogenous cone $\{w|\ w^T A \leq 0, w \geq 0\}$.

These are delivered by restricted F-M elimination. Each extreme ray consists of a set of multipliers (also called weights) for the inputs and outputs. For a particular DMU the 'best' extreme ray(s) provide the maximum possible ratio of outputs to inputs as well as its efficiency number.

The 'best' extreme ray will also give an efficiency number of 1 for certain (efficient) DMUs. These will be the comparator DMUs which we can associate with each extreme ray. From the Duality theorem of LP (see eg. Dantzig(1963)), the variables for these comparator DMUs , in the corresponding Primal model, will be the only ones which are active. In order to find the multiples of the comparator DMUs which show the current DMU to be inefficient, we solve the binding constraints as *equations* for the variables representing the comparator DMUs.

A particular set of comparator DMUs represent the vertices of a region of the efficient frontier in the input-output space. For a model with p inputs and q outputs this frontier will normally have dimension p + q - 2. The ray from the origin to the point representing the DMU under consideration must pass through one such region, the vertices of which represent its comparators. This is illustrated in the next section with a numerical example.

3. A NUMERICAL EXAMPLE

We consider an example, taken from Cooper et al (1999), with 12 DMUs, each of which has 2 inputs and 2 outputs. These are given in Table 1 below.

DMU	1	2	3	4	5	6	7	8	9	10	11	12
Inputs	20	19	25	27	22	55	33	31	30	50	53	38
	151	131	160	168	158	255	235	206	244	268	306	284
Outputs	100	150	160	180	94	230	220	152	190	250	260	250
	90	50	55	72	66	90	88	80	100	100	147	120

Table 1

The Primal model is:

Minimise θ

subject to:
$$-20\,\lambda_1 \; - 19\,\lambda_2 - \ldots \; - 38\,\lambda_{12} \; + x_{1k}\theta \quad \geq 0$$
$$-151\,\lambda_1 \; - 131\,\lambda_2 - \ldots \; - 284\,\lambda_{12} + x_{2k}\theta \quad \geq 0$$
$$100\,\lambda_1 + 150\,\lambda_2 + \ldots + 250\,\lambda_{12} \qquad \geq y_{1k}$$
$$90\,\lambda_1 + \; 50\,\lambda_2 + \ldots + 120\,\lambda_{12} \qquad \geq y_{2k}$$
$$\lambda_1,\,\lambda_2,\ldots.,\,\lambda_{12} \; \geq 0$$

The relevant extreme rays of the cone associated with the corresponding matrix *B*, together with the related (efficient) comparators, are given in table 2 below.

Ray	Extreme Rays (Weights for Inputs and Outputs)	Comparator DMUs
1	(0.808, 0.797, 0.465, 1)	1, 2, 4
2	(0, 1.111, 0.637, 1)	2, 4
3	(0, 0.867, 0.409, 1)	1,4
4	(7.727, 0, 0.645, 1)	1, 2
5	(0, 0.596, 0, 1)	1
6	(4.5, 0, 0, 1)	1
7	(0, 1.145, 1, 0)	2
8	(7.899, 0, 1, 0)	2

Table 2

N.B. Since these vectors represent rays, they can be scaled by any positive multiplier and still represent the same ray. For convenience we have scaled the multiplier for

the second output to be 1, where non-zero, and otherwise scaled the first output multiplier to be 1.

The Efficiency Number of DMUk with inputs x_{1k}, x_{2k} and outputs y_{1k}, y_{2k} is given by:

$$\text{Max } [0.465y_{1k}+y_{2k} \,/\, 0.808x_{1k}+0.797x_{2k};\; 0.637y_{1k}+y_{2k} \,/\, 1.111x_{2k};\; \ldots; \; y_{1k} \,/\, 7.899x_{1k}] \qquad (2)$$

Clearly the multipliers in the eight ratios of (2) are the coefficients from the eight extreme rays above. (Note that the coefficients in the numerator and denominator of any of the ratios in (2) can be scaled by equal amounts). The formula for evaluating the efficiency of any DMUk is contained in (2). One way to elucidate it is to recognise that the homogeneous cone designated as $\{w|\ w^T A \leq 0,\ w \geq 0\}$ in the last section defines the common constraints in the dual LP model. For this example it gives the 24 common constraints on the weights given to inputs and outputs. In the dual model for any of the twelve DMUs these 24 constraints have to be satisfied. Only one constraint and the objective function are distinct for each DMU. The only distinct constraint is: $\sum_i u_i\, x_{ik} = 1$ and the objective is to maximise $\sum_t v_k\, y_{tk}$. Now any linear function can be maximised over a set of linear constraints by evaluating it at all the extreme points of the feasible region. For DMUk if we try and maximise the ratio $\sum_t v_k\, y_{tk} \,/\, \sum_i u_i\, x_{ik}$ where the denominator is restricted to be 1, we can achieve this by evaluating the ratio at all the extreme points of the feasible region for $wA \leq 0$, $w \geq 0$, i.e. at all the rays of the cone, and picking the maximum ratio. Table 3 gives all the eight ratios in (2) for all 12 DMUs.

RAY DMU	1	2	3	4	5	6	7	8	MAX
1	1	0.9161	1.0000	1.0000	1.0000	1.0000	0.5784	0.6333	1
2	1	1	0.9807	1.0000	0.6404	0.5848	1	1	1
3	0.8761	0.8827	0.8685	0.8193	0.5767	0.4889	0.8733	0.8107	0.8827
4	1	1	1	0.9020	0.7190	0.5926	0.9357	0.8444	1
5	0.7635	0.7170	0.7626	0.7451	0.7008	0.6667	0.5196	0.5412	0.7635
6	0.7953	0.8348	0.8328	0.5611	0.5922	0.3636	0.7877	0.5297	0.8348
7	0.8895	0.8738	0.8738	0.9020	0.6283	0.5926	0.8176	0.8444	0.9020
8	0.7963	0.7726	0.7962	0.7435	0.6516	0.5735	0.6444	0.6211	0.7963
9	0.8612	0.8153	0.8402	0.9604	0.6876	0.7407	0.6801	0.8022	0.9604
10	0.8515	0.8706	0.8706	0.6765	0.6260	0.4444	0.8147	0.6333	0.8706
11	0.9345	0.9195	0.9551	0.7687	0.8060	0.6164	0.7420	0.6214	0.9551
12	0.9191	0.8850	0.9028	0.9582	0.7089	0.7018	0.7688	0.8333	0.9582

Table 3

The argument of (2), giving rise to the maximum ratio for a particular DMU, corresponds to the maximum entry in the last column of Table 3 for that row from where the corresponding extreme ray can be worked out. Then the comparators for that ray can be read off from table 2. If the argument of (2) is not unique, (for example it is 2 or 3 for DMU10) multiple sets of weights (and any convex combination of them) are applicable and can be read off from Table 2.

Example 1: DMU 3 has inputs 25 and 160 and outputs 160 and 55. Substituting in (2) the maximum ratio comes from the 2nd argument (ray 2):

$$(0.637*160 + 1*55) / (1.111*160) = 0.883 \tag{3}$$

showing DMU3 to be inefficient with an efficiency score of 0.883. Its comparators are DMU2 and DMU4 taken from Table 2. [Note that the maximum value in Table 3 is .8827 rather than .883 simply because four digit accuracy is used.]

The multipliers 0.637 and 1 for the outputs and 0 and 1.111 for the inputs in (3) are the weightings which DMU3 should choose in order to maximise its ratio of weighted outputs to inputs.

In order to find the quantities of the comparators DMU2 and DMU4 with which DMU3 should be compared, we note that input 2 and outputs 1 and 2 have positive weightings. We therefore solve the corresponding inequalities in the Primal model as equations to give:

$$131\lambda_2 + 168\lambda_4 - 160\theta = 0$$
$$150\lambda_2 + 180\lambda_4 = 160$$
$$50\lambda_2 + 72\lambda_4 = 55$$

leading to $\lambda_2 = 0.9$, $\lambda_4 = 0.139$, $\theta = 0.883$;

ie the artificial DMU made up from these quantities of DMU2 and DMU4 produces the same outputs as DMU3 using no more than 88.3% of all the inputs.

Example 2: We deduce the efficiency of DMU1.

DMU1 has inputs 20 and 151 and outputs 100 and 90. Substituting in (2) the maximum ratio comes from arguments 1,3,4,5 and 6, giving eg

$(0.465*100 + 90) / (0.808*20 + 0.797*151) = 1$ if we use argument 1, and showing DMU1 to be efficient (efficiency number 1). Clearly one of its efficient 'comparators' is itself.

DMU1 could also choose the weightings from any of the rays 3,4,5 and 6 (and any convex combination of them) to maximise its weighted ratio of outputs to inputs.

Example 3: We deduce the efficiency of DMU10.

From table 3 it is clear that DMU10 achieves its maximum ratio from arguments 2 and 3. So its comparators could be DMU2 and DMU4 or DMU1 and DMU4. The primal solution is degenerate, with $\theta = 0.8706$ achieved by letting $\lambda_4 = 25/18 = 1.3889$ and letting all other $\lambda_j = 0$. Using DMU4 at level 1.3889 uses 37.5 of input 1 and 233.33 of input 2 and produces 250 of output 1 and 100 of output 2. So the level of output is the same as DMU10 but only 0.8706 of input 2 is required and even less (the slack is active) of input 1. The dual has multiple optimal solutions – the weights corresponding to ray 2 or ray 3 (and any convex combination of them) are applicable. Note that input 1 has zero weight in both, signifying that both dual solutions are degenerate and indicating the possibility of multiple optimal solutions to the primal as well. In fact, however, there is a unique optimal solution to the primal, representable by many different bases.

Figure 1 gives the skeleton of the (two dimensional) efficiency frontier.

The *efficient* DMUs on this frontier, viz., 1, 2 and 4 are circled and any two of them are connected by a line if they achieve their maximum ratio at a common ray of input and output weights.
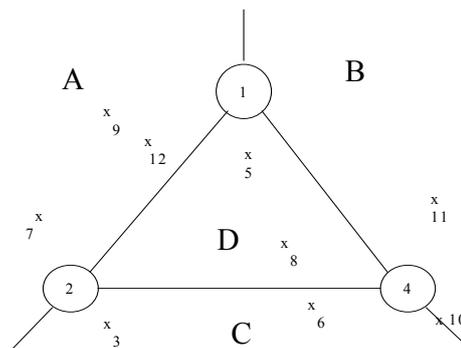

Figure 1

The positions on the skeleton where the rays from the origin to a particular DMU intersect are shown. For example it can be seen that DMU3 intersects the region C showing that the efficient DMUs 2 and 4 are its comparators.

Within each region the same set of multipliers will apply for *all* DMUs whose intersection point lies in this region. The regions are marked and the corresponding input and output multipliers which apply are given below in table 4.

Region	Input Multipliers		Output Multipliers	
A	7.727	0	6.45	1
B	0	0.867	0.409	1
C	0	1.111	0.637	1
D	0.808	0.797	0.465	1

Table 4

In addition DMUs 1 and 2 also each have alternate multipliers arising from the input and output weights of rays 5, 6 and 7, 8 respectively.

Where a DMU's intersection point is on the boundary of different regions the multipliers for *each* region (and any convex combination of them) are alternatives. For DMU10 in example 3 above it was seen that both the multipliers for regions B and C apply. For DMU1 in example 2 the multipliers for regions A, B and D as well as those arising from rays 5 and 6 are applicable. This conforms with the LP result that a degenerate primal solution has alternate corresponding dual solutions. For this model, with 2 inputs and 2 outputs, the primal model will produce at most 3 comparators for any DMU. If there are fewer (as must be the case, for example with an efficient DMU) then the solution is degenerate and alternate sets of multipliers (dual values) are possible.

4. Applications

We now provide some applications to some well known problems considered in the DEA literature.

i)   Returns to Scale

There is a vast literature on the theory and applications (see Banker et al (2004) for a comprehensive survey) of measuring returns to scale (RTS) for DMUs in a DEA model. The standard approach is to declare a DMU as facing increasing (decreasing, constant) returns to scale if the effect of increasing all its inputs by a factor f leads to an increase in all its outputs by a factor greater than (smaller than, equal to) f. The best computational technique for establishing the RTS status of a

DMU, i.e., whether it faces increasing returns to scale (IRS), decreasing returns to scale (DRS) or constant returns to scale (CRS) was published in Zhu and Shen (1995). Their approach is one of four possible approaches contained in a working paper of Appa and Yue (1996) which require a solution of at most two LPs per DMU to resolve the RTS status of a DMU. We now show that the information obtained through only one pass of the F-M method is sufficient to establish the RTS status of all the DMUs.

For each DMU the F-M method delivers all the possible sets of weights for which the argument in (1) is maximum. We have illustrated this for the CCR model of DEA. As shown in the numerical example of section 3, calculating the λ_j values corresponding to a particular set of optimal weights is a trivial exercise consisting of solving a system of r (= p+q+1) independent linear equations in r variables. Let us denote the optimal values of λ_j by λ_j^*. Then the value of $\Sigma\lambda_j^*$ corresponding to each optimal set of weights is also available after one pass of the F-M method.

Now we use the analysis developed in Zhu and Shen (1995) to show that if we have all the optimal weights and the corresponding optimal $\Sigma\lambda_j^*$ values from the CCR model, the RTS status of every DMU is immediate. First, note that all efficient DMUs in the CCR model face CRS and have at least one solution with $\Sigma\lambda_j^* = 1$. Further, as shown in a theorem proved in Zhu and Shen (1995):

Theorem: DMU_0 faces IRS (DRS) if and only if optimal $\Sigma\lambda_j^*$ is less than (greater than) one in *all* optimal solutions.

It follows that from the output of the F-M model we have all the necessary ingredients to decide the RTS status of all the DMUs. This is in contrast with the normal procedure for solving DEA models where at least one more LP will have to be solved for some DMUs to determine whether in *all* optimal solutions $\Sigma\lambda_j^* < 1 \, (> 1)$ or not. We illustrate with the numerical example in section 3.

DMUs 1, 2 and 4 are efficient and therefore face CRS. For DMU3 we have shown in section 3 that the optimal values of non-zero λ_j's are $\lambda_2^* = 0.9$ and $\lambda_4^* = 0.139$. So $\Sigma\lambda_j^* = 1.039$ and DMU3 faces DRS. The case of DMU10 is particularly interesting as there are two distinct optimal weight sets. For both $\Sigma\lambda_j^* = 1.3889$, so DRS prevails for DMU10 as well. A similar analysis gives the RTS status of all the DMUs summarised in the table below.

DMU	1	2	3	4	5	6	7	8	9	10	11	12

RTS status	CRS	CRS	DRS	CRS	IRS	DRS	DRS	DRS	DRS	DRS	DRS	DRS

Table 5 RTS status of 12 DMUs from one pass of the F-M method

ii) Cross-efficiency

The standard approach in DEA allows each DMU to appear in the best possible light by selecting its own weights. This self-appraisal based simple efficiency can be contrasted with cross-efficiency, calculated on the basis of peer-appraisal. The normal method of doing this (see Doyle and Green (1994), Sexton (1986)) selects a set of weights for each DMU that maximises the ratio of its weighted sum of outputs to weighted sum of inputs and then evaluates all other DMUs at these weights as well.

Formally, let the vector $w^h = \{u^h_{ik}, v^h_{tk}\}$ represent a set of optimal input and output weights for DMUk corresponding to θ^* for some argument m of (1). Then the cross-efficiency score of DMUs ($s \neq k$), evaluated at any optimal weight set w^h is denoted by E_{ks} and defined as:

$$E_{ks} = \left(\sum_t v^h_{tk} y_{ts} \right) / \left(\sum_i u^h_{ik} x_{is} \right) \tag{2}$$

The following diagram illustrates the concept.

13
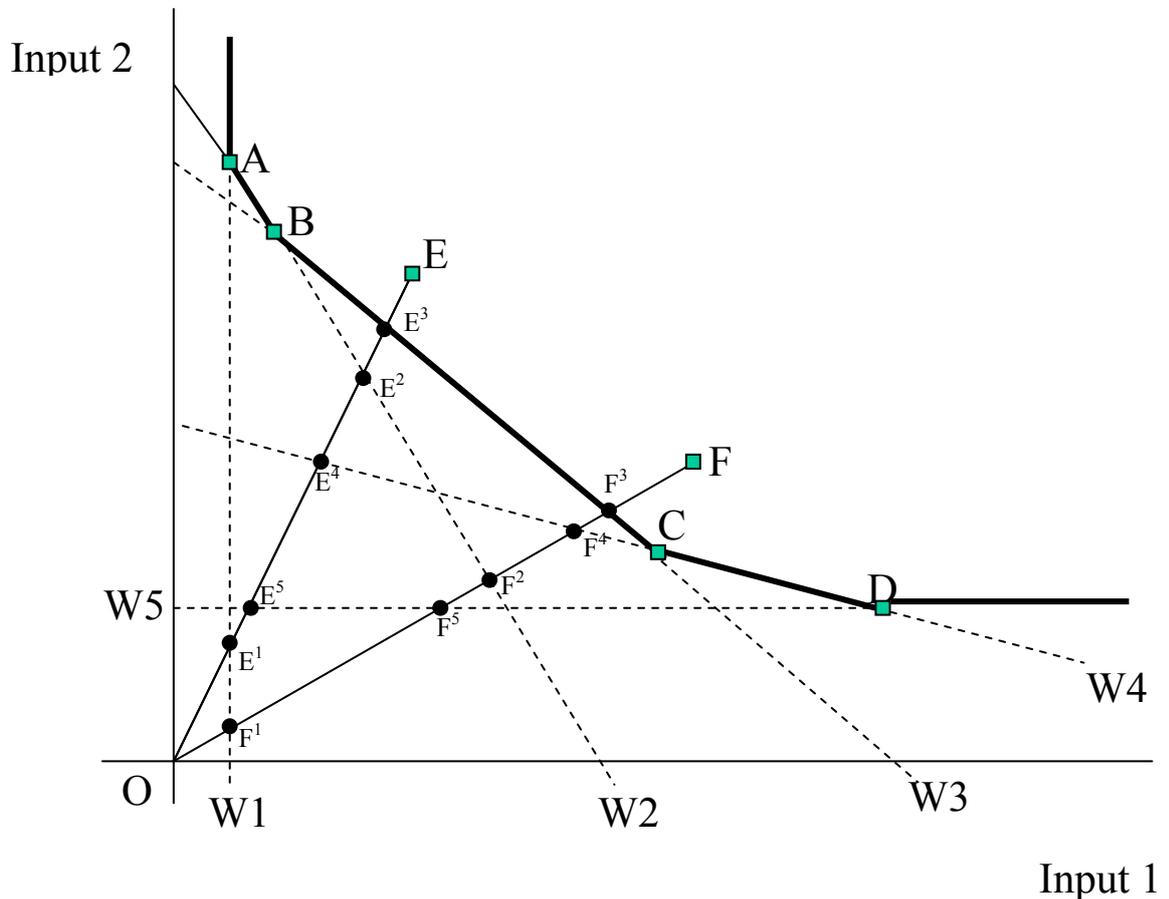
The CRS efficiency frontier for this standard 2-inputs, 1-output diagram is given by ABCD adjoined by the vertical line above A and the horizontal line to the right of D. The complete set of optimal weights of inputs 1 and 2, corresponding to the rays obtained from a pass of the F-M method, are given by the normals to the broken lines W1 to W5, when the output weight has been standardised at value 1. We denote these as w^1 to w^5. DMUs A to F are represented by a square and show the quantities of the two inputs needed to produce one unit of the output. All the four CRS efficient DMUs have two distinct sets of optimal weights. For example, both w^3 and w^4 are optimal set of weights for DMU$_C$. Cross-efficiencies are illustrated for DMUs E and F. The cross-efficiency of DMU$_E$ at weight set Wg is OE^g/OE. Similarly, OF^g/OF is the cross-efficiency of DMU$_F$ for g = 1 to 5.

The computation of E_{ks}, however, is complicated by the presence of multiple sets of optimal weights for DMUk because the value of E_{ks} is not unique. For example, E_{CE}, the cross-efficiency of DMU$_E$ evaluated at an optimal set of weights for DMU$_C$ is higher if our choice of optimal weight set for DMU$_C$ is w^3 rather than w^4. Moreover, if w^{g1} and w^{g2} are two optimal weight sets for a DMU, it cannot be guaranteed that *all* other DMUs will be evaluated at a higher cross-efficiency by w^{g1} as compared to w^{g2}. For example, consider E_{CF} and E_{CE} in the diagram. E_{CF} is higher for DMU$_F$ if weight set w^4 rather than w^3 is used for DMU$_C$, but vice versa for E_{CE}.

To define E_{ks} uniquely, Sexton et al (1986) and Doyle and Green (1994) employ the concept of an aggressive or a benevolent strategy. In the aggressive strategy DMUk will choose a set of weights that minimises the average of the n-1 other DMUs' cross-efficiencies when rating them by applying DMUk's best weight. So, given H best or optimal weights for DMUk denoted by $w^h = \{u^h_{ik} , v^h_{tk}\}$ for h = 1 to H, DMUk should select the set $w^{h*} = \{u^{h*}_{ik} , v^{h*}_{tk}\}$ which minimises (3) below over h.

$$\frac{1}{n-1}\sum_{s\neq k} E^h_{ks} = \frac{1}{n-1}\sum_{s\neq k} \left(\sum_t v^h_{tk} c_{ts}\right) / \left(\sum_i u^h_{ik} a_{is}\right) \tag{3}$$

If a benevolent strategy is employed, the expression in (3) is maximised over h instead. But neither Sexton et al (1986), nor Doyle and Green (1994), nor any other authors to our knowledge, are able to find a weight set that achieves the minimum or the maximum of the non-linear fractional programming problem implied by (3). All the contributors to the field seem to use a surrogate measure for (3) whereby the

inputs and outputs of all DMUs other than k are summed together to create one composite DMU to which the P sets of optimal weights for DMUk are applied.

Our method does away with the need to create surrogate measures. We can actually compute the minimum or maximum of (3) from the output of the F-M method. For example, table 3 gives all the ratios required in (3) for each weight set. So, for DMU1 of the 12 DMU example in section 3, the total number of optimal weight sets H = 5, corresponding to ray 1 and rays 3 to 6. We calculate the average of cross-efficiencies – call it AE_1 - for each one of these weights by adding the ratios in rows 2 to 12 of table 3 and dividing the sum by 11. This gives:

w^h	w^1	w^3	w^4	w^5	w^6
Average cross-efficiency	0.880636	0.8803	0.821527	0.667045	0.578727

Table 6 Direct calculation of average cross-efficiency without surrogates

It is obvious that to implement the benevolent strategy we should choose w^1 and for the aggressive strategy choose w^6 for DMU1. So there is no need to find surrogates.

Having shown how the F-M method improves the calculation of cross-efficiency, we point out two further shortcomings of the usual surrogate models that can be tackled by our method. First, notice that the solution to (3) may not be unique. For example, there is no significant difference between the AE_1 values for optimal weight sets w^1 and w^3 in the table above and both of these could be regarded as giving the maximum average. This can also happen in the surrogate models solved by Doyle and Green (1994) and others. Secondly, the weight set selected for the aggressive strategy may itself be a 'maverick' weight set in the sense that it gives non-zero values to very few inputs and outputs. Our selection of w^6 for DMU1 under aggressive strategy illustrates the point. How to rectify these shortcomings is an obvious topic for further research. An obvious way forward would seem to be the introduction of further objectives, over and above the finding of a weight set for DMUk that maximises its own efficiency and minimises the average efficiency of its peers.

iii) Weight restrictions

An obvious area of application is to models attempting to restrict weights in some ways. These models look at the feasible region of weights and introduce additional constraints to overcome the problems introduced by complete weight

flexibility in DEA. Thompson et al (1986) introduced upper and lower bounds on the ratio of some selected pairs of weights, calling the restricted feasible region the Assurance Region (AR). In the so called Polyhedral Cone-Ratio (PCR) model, Charnes et al (1989, 1990) select some exemplary efficient DMUs from among all the efficient ones and construct the convex cone of the weights used by these efficient DMUs as the feasible region of weights. A common problem for all these models is that they work with only those weights revealed by solving n LPs for establishing the efficiency of n DMUs. However, a fail-safe application of the models requires a knowledge of the entire feasible region of weights. Our method can be easily extended to find the efficiency levels and all the optimal set of weights for all the DMUs within the restricted feasible region enforced by either AR or PCR models because both of them merely impose further restrictions on the weights w of the type $wD \leq 0$. Together with the original set of inequalities, these still define a convex cone. So the same procedure will find all the necessary information. But it delivers much more. We illustrate below how our method helps in the selection of exemplary DMUs for the PCR model

Tone (1997) attempted to use a similar approach to ours to find solutions to a CCR DEA model with PCR restrictions on weights. Motivated by the difficulty of finding the complete set of weights used by the exemplary efficient DMUs, he uses the double description method – a version of the F-M method - for finding all the extreme points of a polyhedral region. In the various versions of the PCR model that he considers, he solves n LPs to find the set of efficient DMUs first and then uses the double description method either to find the complete set of optimal weights for each exemplary DMU or to find the set of inequalities $wB \leq 0$ for a given sub-set of weights. We show below how, without solving n LPs or doing any extra work at all, our method can help in the selection of exemplary DMUs by providing a range within which efficiency levels of every DMU will fall after enforcing weight restrictions based on exemplary DMUs.

First of all, note that under weight restrictions of this type there is always a feasible solution, because the optimal weights of the exemplary DMUs are a feasible set of weights that can be used by all the other DMUs. If the set of optimal weights for the exemplary DMUs is denoted by W^E, the maximum ratio of

$$\left[\left(\sum_{t=1}^{q} v_{tl} \; y_{tk}\right) \bigg/ \left(\sum_{i=1}^{p} u_{il} \; x_{ik}\right)\right]$$ for $l \varepsilon W^E$ gives a lower bound for the new efficiency level

for DMUk. This information is not available from the conventional approach to solving DEA models. Further, the efficiency rating of a DMU cannot increase under weight restrictions. So the unrestricted maximum ratios provide upper bounds. Together these lower and upper bounds provide a powerful tool for choosing amongst alternative sets of exemplary DMUs.

By way of illustration, consider the numerical example of section 3. Suppose we show the results given in table 2 to our expert who recommends that we use DMUs 1 and 4 as exemplary DMUs. Then we are concerned with re-evaluating the efficiency of all the DMUs using only weights which are in the convex cone of rays 1 to 6, ignoring the rays 7 and 8. We now wish to re-evaluate all the 12 DMUs under weight restrictions limiting the weights to be in the convex cone of w^1 to w^6. An elementary grasp of convex cones tells us that rays 1 to 6 will continue to be the extreme rays of the restricted cone, while some new extreme rays may also emerge. So we already have a lower bound on the efficiency levels of each DMU given by the maximum ratio in columns 1 to 6 in table 3. For our specific example, it can be easily checked that these lower bounds coincide with the current best for each DMU. So the weight restrictions make no difference to efficiency scores if the PCR model is defined as above. A closer scrutiny of the Table 3 reveals that selecting any two of the three efficient DMUs as exemplary DMUs will lead to the same impasse simply because all the DMUs achieve their maximum ratio on at least one of the first four rays. So the PCR model will give different efficiency scores provided only one efficient DMU is taken as an exemplary DMU.

Suppose DMU2 is considered to be the only exemplary DMU. So $W^E = \{w^1,$ w^2, w^4, w^7 and $w^8\}$. From Table 3 we can see that this would lead to a reduction in the efficiency score of DMU 11 only as the weights defined by ray 3, the one giving its maximum ratio, are no longer available. DMU 10 is not affected because it achieves its maximum ratio on Rays 2 and 3. Table 3 informs us that for DMU11 the efficiency score will be between 0.9345 (the best achieved on rays being considered for this PCR model) and 0.9551, its best score without any weight restrictions. It is also clear from Table 3 that no matter which efficient DMUs are selected as exemplary in a PCR model for this problem, the efficiency score of DMUs 1, 2, 4, 5,

8 and 10 will not be affected. The following table gives the impact of selecting any one of the three efficient DMUs as exemplary.

Exemplary DMU	$W^E =$	Range of Efficiency values θ^k for affected DMUk
1	$\{w^1, w^3, w^4, w^5, w^6\}$	$.8761 \leq \theta^3 \leq .8827;\ .8328 \leq \theta^6 \leq .8348$
2	$\{w^1, w^2, w^4, w^7, w^8\}$	$.9345 \leq \theta^{11} \leq .9551$
3	$\{w^1, w^2, w^3\}$	$.8895 \leq \theta^7 \leq .9020;\ .8612 \leq \theta^9 \leq .9604;$ $.9191 \leq \theta^{12} \leq .9582$

Table 7 Impact of alternative PCR models

What we have shown is that using the F-M method provides a powerful tool for the selection of the right PCR model by providing us with the range within which efficiency of each DMU can be expected to change. Similar approach can be used in all models of weight restrictions.

iv) Sensitivity Analysis

Another obvious area of application is sensitivity analysis, in particular the effect of data variations on efficiency of DMUs. Using the richness of information available in tables 2 and 3, it is easy to analyse the range of values of an input (output) x_{ik} (y_{tk}) for which the maximum ratio in (2) is derived from the same ray. This is obviously a rich field for further research.

To conclude this section we observe that our method can help with any DEA related problem where all possible optimal solutions to the primal or the dual DEA model are required for proper implementation.

5. Computational considerations and Conclusions

We begin this section with a note on computations required to solve DEA problems within the new framework. Standard packages are available in the public domain to carry out the calculations needed to implement the F-M method. The calculations for all the computational experiments reported here were carried out by PORTA, a general purpose code for finding all the extreme rays of a convex cone. PORTA Version 2, available from Free Software Foundation, Boston, 1991 Program was interfaced to our own C program.

We have provided an alternative framework for solving DEA problems. The purpose is *not* to provide a faster alternative to the conventional LP based approach,

but to use a method that delivers more. In fact the method used is known to take longer than the conventional method because the projection of a polytope into a smaller dimension can lead to an exponential growth in the number of in the number of inequalities generated in finding the extreme rays. We carried out computational experiments to determine if this is likely to be an impediment to solving large DEA models in this way. The data sets available at http://www.mscmga.ms.ic.ac.uk/jeb/orlib/deainfo.html and http://www.deazone.com/datasets/FILE1/index.asp were utilised. No serious computational problems were encountered.

We also used our approach in a practical problem involving the computation of the x-factor in electricity price regulation in Brazil, based on measuring technical efficiency of 63 electricity distribution companies. (See Appa and Bana-e-Costa, (2003)). A complex model involving dynamic clusters of companies taken as comparators for each company was implemented on a one input, four output CRS DEA model using the F-M method. It was easy to compute results in real time and use them as an interactive tool in deciding the best parameters on which the dynamic clusters should be based.

The lessons learnt are:

(i)   As the problem size grows, performance measures such as number of rays, maximum no. of inequalities produced at any iteration, sum of inequalities dealt with, maximum memory used, operating time etc. grow rapidly.

(ii)   The effect of increasing number of factors (inputs + outputs) is much greater than that of increasing the number of DMUs.

(iii)   A problem with only one input or only one output is easier to solve than a comparable one with same number of factors but multiple inputs as well as outputs.

(iv)   For problems tried in our experiments (with up to 70 DMUs and 11 factors) the computational growth is easily manageable. Interactive use of the methodology on a practical problem does provide further evidence of its manageability.

While more systematic work is needed to check the computational limits of the F-M method applied to DEA, based on our experience so far we maintain that the method proposed here is a viable method since:

(i)   Most practical DEA models tend to be small in comparison with the large LP models that arise in other applications.

(ii)   In practical DEA problems the number of factors is usually small.

Finally, there are two obvious ways, currently under investigation, to control the growth in the number of inequalities created to find the extreme rays of the cone. These are:

(i)   To exploit the structure of the *A* matrix in DEA models. The DEA models have a special structure with positive coefficients in one of the row partitions and negative coefficients in the other. Experience with F-M elimination suggests this will reduce the growth in inequalities when projecting out variables.

(ii)   To control the growth of inequalities solved to compute extreme rays by a two phase approach. In the first phase use the standard approach to determine the set E of efficient DMUs. Then run the F-M method only on the restricted set E. As the extreme rays are determined only by the efficient DMUs, this helps to control the problem of exponential growth.

For this preliminary exposition we have deliberately kept things simple by looking at the original constant return to scale CCR model. We are aware that the standard constant returns to scale model in modern literature minimises $w + \varepsilon(T)$ where T is the sum of all the slack and surplus variables to deal with points on the frontier which have $w = 1$ but $T \neq 0$. Other models such as the variable returns to scale model, models with restrictions on weights etc. are also of great interest for practical work. We have been able to solve all these models using the Restricted Fourier – Motzkin Elimination method.

Acknowledgments

REFERENCES

Appa, G., Bana-e-Costa, C., 2003.A Methodology to estimate the relative operational performance of electricity distributors in Brazil *LSE OR department Working Pape Seriesr*. LSEOR 03-57, the ISBN number 0 7530 1521 8.

Appa.G, Yue, M. 1996. |On Measuring Scale Efficiency and Setting Scale Targets in DEA. *LSE OR department Working Paper Series*. LSEOR 96.22/2.

Banker, R., Cooper W.W., Seiford, L.M., Thrall, R.M. and Zhu, J. 2004. Returns to scale in different DEA models. *Eeuropean Journal of Operational Research*. 154, 345-362.

Charnes, A., Cooper, W.W., Rhodes, E. 1978. Measuring the Efficiency of Decision Making Units. *European Journal of Oper. Res*. 2, 429-444.

Cooper, W. W., Seiford, L.W., Tone, K. 1999. *Data Envelopment Analysis*. Kluwer Academic Publishers.

Dantzig, G.B., 1963. *Linear Programming and Extensions,* Princeton University Press, Princeton, New Jersey

Doyle, J., Green, R., 1994. Efficiency and Cross-efficiency in DEA. *JORS* 45, 567-578.

Harris II, J., Ozgen, H. and Oscan, Y., 2000. Do Mergers Enhance the Performance of Hospital Efficiency? *JORS* **51,** 801-11

Sexton, T.R., Silkman, R.H. and Hoogan, A.J., 1986. Data Envelopment Analysis: critique and extensions. In *Measuring Efficiency: An Assessment of Data Envelopment Analysi*. (R. H. Silkman, Ed.) 73-104.

Seiford, L.M. 1995. *A Bibliography of Data Envelopment Analysis (1978-1995)* Thanassoulis, E., 2001. *Introduction to the Theory and Application of Data Envelopment Analysis*, Kluwer Academic Publishers

Seiford, L. M. and Zhu, J.,(1999). "An Investigation of Returns to Scale in Data envelopment Analysis". *Omega*, **27**, 1-11.

Tone, K. (1997) Several Algorithms to Determine Multipliers for Use in the Cone-ratio Envelompent Approaches to Efficiency Evaluation in DEA, *Computational Approaches to Economic Problems*, (Amman, Rustem and Whinston, eds, Kluwer pp. 91-109

Williams, H.P. (1986), Fourier's Method of Linear Programming and its Dual, *Am.Math Monthly* **93,** 681-94

Zhu, J. and Shen, Z. H., A discussion of testing DMU's returns to scale. *European Journal of Operational Research* 1995, **81**, 590-596.