# A methodology for cross-evaluation in DEA

G. Appa[1] , N. Argyris[1] and H.P. Williams

{G.Appa; N.Argyris; H.P.Williams}@lse.ac.uk

Operational Research Group, Department of Management,

London School of Economics and Political Science

Houghton Street, London WC2A 2AE, UK

Working Paper No: LSEOR 06.81

November 7, 2006

[1] The authors' research was supported by the UK Government Department for Education and Skills, Grant No. EOR/SBU/2003/208

First published in Great Britain in 2006 by the Operational Research Group, Department of Management, London School of Economics and Political Science







Typeset, printed and bound by:

The London School of Economics and Political Science

Houghton Street

London WC2A 2AE

Working Paper No: LSEOR 06.81

ISBN No: 07530 1703 2

**Abstract**

This paper provides a new approach for cross-evaluation in DEA. An examination of the link between cross-evaluation and the DEA production possibility set reveals flaws in the current methodology for cross-evaluation. This is compounded by problematic cases permitted by the existing approaches for performing cross-evaluation without recourse to the full set of optimal weights for all DMUs. Our new approach is based on information on all weighting schemes for all DMUs. We show how this overcomes the existing problems. We also provide new tools for the identification of maverick DMUs that make use of unrealisitic weighting schemes and introduce the concept of under-achieving DMUs.

**keywords:** Data Envelopment Analysis, Cross-Evaluation, Cross-Efficiency, Ranking, Maverick DMUs, Under-achieving DMUs, Weight-Restrictions, Efficient frontier identification

# 1 Introduction

Data Envelopment Analysis (DEA) is a widespread technique that evaluates the efficiency of a set of homogeneous production units, termed Decision Making Units (DMUs), that operate in a multiple input and output environment. A part of what can perhaps be called the 'standard' methodology in DEA (first enunciated by Farrell (1957) and then developed by Charnes *et al* (1978) (CCR) and extended by Banker *et al* (1984) (BCC)), involves the use of a ratio of the form: weighted sum of outputs/ weighted sum of inputs, to assign an efficiency score to every DMU. The weighted sums are obtained after assigning multipliers, or weights, to the inputs and outputs. DEA allows each DMU to choose its own multipliers such that its efficiency score is maximised (subject to some constraints). We will refer to the efficiency scores calculated by this self-appraisal as simple efficiencies. Sexton *et al* (1986) proposed that the optimal weights for each DMU can be used to appraise its peers, i.e. to calculate alternative efficiency scores for every other DMU. We refer to the efficiency scores calculated in such a way as cross-efficiencies and use the term cross-evaluation (Doyle & Green, 1995) for the process of evaluating a DMU's cross-efficiency scores. The use of Cross-evaluation has spread to a number of different areas. It can be used in Multiple Criteria Decision Making (MCDM) to improve discrimination among alternatives (see (Doyle, 1995),(Green & Doyle, 1995),(Sarkis, 2000),(Mavrotas & Trifillis, 2006)), in ranking candidates in a preferential election (Green *et al.* , 1996), in the ranking and selection of projects and technologies ((Oral *et al.* , 1991),(Green *et al.* , 1996),(Shang & Shueyoshi, 1995)). Recently, Gregoriou *et al* (2005) used cross-evaluation within a DEA methodology to evaluate the performance of hedge-fund classifications. Other applications include (Sarkis & S.Talluri, 2004),(Chen, 2002),(Ertay & Ruan, 2005). In addition, this concept has real appeal for central organisations in charge of aspects of funding DMUs. For example the UK government's Department for Education and Skills (DfES) are considering utilising cross evaluation as part of their performance measurement methodology for schools. It is planned that such measures will subsequently be placed on a benchmarking website which will assist the dissemination of best practice between schools. (DfES, 2005). Despite the many interesting uses of cross-evaluation, its theoretical framework has been a relatively understudied concept. To the best of our knowledge, apart from the initial theoretical contributions ((Sexton *et al.* , 1986), (Doyle & Green, 1994), (Doyle & Green, 1995)), the only recent development is by Anderson *et al* (2002) who examine the characteristics of the special case of single output (input) models.

Recently the DEA literature has seen the publication of some very interesting research on methods of explicitly identifying the DEA efficient frontier, or equivalently the set of all possible optimal input and output weights (see e.g. (Raty, 2002), (Briec & Leleu, 2003), (Olesen & Petersen, 2003), (Appa & Williams, 2006)).

In this paper we carry out a critical appraisal of both the theoretical framework and the computational tools used in the traditional approaches to cross evaluation. Our work reveals many flaws and weaknesses. These are addressed by providing a novel approach to cross evaluation based on the explicit identification of the efficient frontier. Working with the complete set of optimal weights we compute cross-efficiency scores of each DMU at each possible optimal weighting scheme, leading to a more meaningful ranking of all DMUs, new indicators for maverick weights and DMUs and the identification of under-achieving DMUs.

The next section introduces cross-evaluation and defines the notation used in the rest of the paper. In section three we show that for constant and variable returns-to-scale models (CRS and VRS) cross-evaluation uses points outside the DEA production possibility set. Additionally, for the VRS case, it is shown that there is a contradiction between the cross-evaluation methodology and the philosophy of the VRS model. In section four we identify some of the conceptual and computational weaknesses of the traditional approaches for cross-evaluation. Of necessity, the traditional analysis is incomplete and the set of optimal weights and the scores obtained are not always consistent.

Section five gives our novel approach for cross-evaluation that addresses the problems identified in earlier sections. This leads to consistent average cross-efficiency scores and ranking of DMUs. Additionally, it enhances the accuracy of the traditional maverick index and leads to the introduction of a new indicator that allows for the identification of DMUs using 'unrealistic' weighting schemes in a coherent manner. When the new indicator takes negative values, we are alerted to a hitherto unnoticed phenomenon of an under-achieving DMU which compares poorly with the average of cross-efficiency scores of others at its favourite weighting scheme. We also show how information on all weighting schemes can help when introducing weight restrictions. All the new concepts are illustrated and computational aspects discussed before the final concluding section.

## 2 Cross-Efficiency

The notation we will use is as follows. Let $\mathcal{J} = \{1, 2, ..., n\}$ be a set of DMUs indexed by $j$. DMU $j$ uses the $s$-dimensional input vector $\mathbf{x}_j$ to produce the $m$-dimensional output vector $\mathbf{y}_j$. We also define the matrices $\mathbf{X} = [\mathbf{x}_1^T, ..., \mathbf{x}_n^T]$ and $\mathbf{Y} = [\mathbf{y}_1^T, ..., \mathbf{y}_1^T]$. Finally, let $v \in \mathbb{R}_+^s$ and $u \in \mathbb{R}_+^m$ be the vectors of weights that DMUs assign to their inputs and outputs respectively. To properly illustrate cross-evaluation we will use the following non-linear programme by CCR (1978), which evaluates the efficiency of DMU $O \in \mathcal{J}$:

$$\max \frac{u\mathbf{y}_o}{v\mathbf{x}_o} \tag{1a}$$

$$s.t. \ \frac{u\mathbf{y}_j}{v\mathbf{x}_j} \leq 1 \ \forall j \in \mathcal{J} \tag{1b}$$

$$u, v \geq 0 \tag{1c}$$

Note that for simplicity we use $u$, $v$ instead of $u^T$, $v^T$. The optimal solution to (1), denoted $(u^*, v^*)$, provides a set of weights (or a weighting scheme) that maximises the ratio in the objective function, i.e. the efficiency of $\text{DMU}_O$. It is worth noting however that the model requires that these weights do not yield efficiency scores greater than one when applied to any DMU. In other words, all cross-efficiencies are constrained to be less than or equal to one. We can now formally define cross-efficiency:

**Definition 1** *Let $(u_k^*, v_k^*)$ denote a set of optimal weights for DMU $k$, obtained by solving model (1). The cross-efficiency score of DMU $j$ relative to DMU $k$, denoted $h_{jk}$, is given by:*

$$h_{jk} = \frac{u_k^*\mathbf{y}_j}{v_k^*\mathbf{x}_j} = \frac{O_{jk}}{I_{jk}} \tag{2}$$

*where $O_{jk}$ $(I_{jk})$ is the value of DMU $j$'s outputs (inputs) evaluated by applying weights $u_k^*$ $(v_k^*)$.*

This directly implies that at the optimal solution for (1), $\frac{u_k^*\mathbf{y}_j}{v_k^*\mathbf{x}_j}$ - the value of the left hand side of every constraint in (1b), gives the cross-efficiency score for a DMU (including the simple efficiency for $\text{DMU}_O$). For simplicity, we will denote the simple efficiency of DMU $j$ by $h_j$ instead of $h_{jj}$. The information given by all cross-evaluations can be most conveniently summarised in the $n \times n$ cross-efficiency matrix given in Table 1. The entries of the leading diagonal are the simple efiiciencies. One can average across rows to obtain the average appraisal of peers by DMU $k$ (denoted $A_k$), or across columns to obtain the average appraisal of DMU $j$ by peers (denoted $e_j$). We will focus on the second measure, taken to be the arithmetic mean of cross-efficiency scores, and refer to it as *average cross-efficiency*. Instead of the mean, one could also consider the median or variance of the cross-efficiency scores (Doyle & Green, 1994). Averaging can be done with or without the leading diagonal, depending on whether we want to include or exclude the self-appraisal in the average score. We choose to include it and define $e_j$ as:

$$e_j = \frac{1}{n}\sum_{k=1}^{n} h_{jk} \tag{3}$$

An advantage of average cross-efficiency over simple efficiency is that the former can be used to rank all DMUs (Doyle & Green, 1995). Traditional DEA does not provide a way to rank efficient units since they all have equal efficiency scores. In cross-evaluation although an average crosss-efficiency score of one (or 100%) is theoretically possible, in practice this is very unusual and would require a strange data-set. In general, the possibility of ties is very low and a unique ranking of all DMUs can be determined (for a review of ranking methods in DEA refer to Angulo-Meza and Lins (2002) and Adler *et al* (2002)).

Another advantage of cross-evaluation is its ability to reduce the effect of unrealistic, or maverick, weighting schemes used by some DMUs. DEA allows for complete weight flexibility, i.e. every DMU is free to choose the weights with which its efficiency will be evaluated. This is an important advantage of DEA to methods that rely on the expertise of the evaluator to identify weighting schemes with which DMUs will be scored. It does however give rise to the following problem: some DMUs might select weighting schemes that will be 'unrealistic' under some managerial criteria. A special case of this behaviour appears when DMUs that make efficient use of only a subset of inputs and outputs assign zero weights to all other inputs and outputs, thereby obtaining 'unrealistic' efficiency scores. The case of zero weights is of special theoretical importance to DEA given its link with weak efficiency. Olesen and Petersen (1996) examine the problems arising by the use of zero weights and point out the importance of using weighting schemes with strictly positive weights for all inputs and outputs which correspond to fully dimensional efficient facets (FDEFs). We would like to point out however that

| Rated DMUs | Rating DMUs | | | | | $e_j$ |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | ... | $n$ | |
| 1 | $h_{11}$ | $h_{12}$ | $h_{13}$ | ... | $h_{1n}$ | $e_1$ |
| 2 | $h_{21}$ | $h_{22}$ | $h_{23}$ | ... | $h_{2n}$ | $e_2$ |
| 3 | $h_{31}$ | $h_{32}$ | $h_{33}$ | ... | $h_{3n}$ | $e_3$ |
| ... | ... | .. | ... | ... | ... | ... |
| $n$ | $h_{n1}$ | $h_{n2}$ | $h_{n3}$ | ... | $h_{nn}$ | $e_n$ |
| $A_k$ | $A_1$ | $A_2$ | $A_3$ | ... | $A_n$ | |

Table 1: Cross-Efficiency Matrix

fully dimensional weighting schemes could also be unrealistic. One way to prevent 'maverick' DMUs from using unrealistic weighting schemes is imposing additional constraints on the input and output weights (weight restrictions) (see e.g. (Thompson *et al.* , 1986)(Dyson & Thanassoulis, 1988)(Charnes *et al.* , 1990)). However, this often requires a priori information and relies on the "expert knowledge of the modeller to create these restrictions" (Anderson *et al.* , 2002)), and such constraints "also have something of the arbitrariness (authoritarianism) that exponents of DEA may have wished to escape" ((Doyle & Green, 1994)). Cross-evaluation reduces the effect of unrealistic weighting schemes since cross-efficiency values are a result of all available weighting schemes. "Rather than have an external weight restriction applied by an expert, the data-set serves as the arbiter of good judgement by, in essence, creating its own weight restrictions" (Anderson *et al.* , 2002).

Doyle and Green (1994) propose the use of the *maverick index* to identify DMUs that select unrealistic weighting schemes. These are also called maverick DMUs. The maverick index measures the relative increment when shifting from average cross-efficiency to simple efficiency and is defined as:

$$M_j = \frac{h_j - e_j}{e_j} \tag{4}$$

A relatively high value of $M_j$ suggests that DMU $j$ is probably using unrealistic weighting schemes for its self-appraisal. In addition, we can use $M_j$ to identify all-round performers. These will be the DMUs with relatively low maverick indices (see (Chen, 2002) and (Ertay & Ruan, 2005) for applications of the maverick index).

# 3 A critical appraisal of the theoretical framework for cross-evaluation

To date, cross-evaluation approaches have not taken into account the attributes of the DEA production possibility set. In this section we examine the link between these two concepts for both constant and variable returns-to-scale models (CRS and VRS) and throw light on some serious problems in providing a consistent theoretical framework to integrate them.

The production possibility set $T$ includes all feasible input and output vectors. The CCR model estimates $T_{CRS}$ by the polyhedral empirical production possibility set $\hat{T}_C$ defined as:

$$\hat{T}_C \equiv \{(x, y) \in \mathbb{R}_+^{s+m} |\ x \geq \mathbf{X}\lambda,\ y \leq \mathbf{Y}\lambda,\ \lambda \geq 0\} \tag{5}$$

Let $\mathcal{E}^C \subseteq \mathcal{J}$ be the set of CCR-efficient DMUs and $P_C$ the polyhedral cone of feasible input-output multipliers corresponding to $\hat{T}_C$:

$$P_C \equiv \{(u, -v) |\ u\mathbf{y}_j - v\mathbf{x}_j \leq 0,\ j \in \mathcal{E}^C,\ v \in \mathbb{R}_+^s, u \in \mathbb{R}_+^m,\ u, v \neq 0\} \tag{6}$$

We can now represent $\hat{T}_C$ in terms of $P_C$ as follows (see (Olesen & Petersen, 1996)):

$$\hat{T}_C = \{(x, y) \in \mathbb{R}_+^{s+m} | \forall (u, -v) \in P_C :\ uy - vx \leq 0\} \tag{7}$$

In simple words, the polyhedral set $\hat{T}_C$ can be characterised by a finite collection of hyperplanes of the form $uy - vx = 0$ (see e.g. (Ali & Seiford, 1993) and (Sueyoshi, 1999)).

We illustrate cross-evaluation with a CRS example given in Figure 1. The DEA production possibility set (input set) is the intersection of five halfspaces defined by the hyperplanes denoted $f_i$ (for $i = 1$ to 5), each corresponding to a weighting scheme given by the normal to the hyperplane. The cross-efficiencies of DMUs are calculated by radially projecting them onto the hyperplanes. For example, the cross-efficiency of DMU E at weighting schemes 1 and 2 are the relative amounts of radial reduction in the input levels of DMU E needed for E to meet hyperplanes $f_1$ and $f_2$ at points $E_1$ and $E_2$ respectively, so that their actual cross-efficiencies are given by the ratios $\frac{OE_1}{OE}$ and $\frac{OE_2}{OE}$.

DMUs A, B, C and D are DEA efficient, each at two sets of weights. DMU F is not efficient but also achieves its optimal efficiency score at two sets of weights. The existence of more than one optimal weighting schemes for DMUs creates problems for cross-evaluation. We explore these in the next section. We would like to stress here that reference points $E_1$, $E_2$, $E_4$ and $E_5$ lie outside the production possibility set. In general, cross-evaluation implies the scoring of DMUs relevant to points that lie outside the DEA production possibility set. In the DEA literature a similar situation arises in methodologies that seek to avoid the use of non-fully-dimensional weighting schemes by controlling the projection of DMUs on the frontier (see (Bessent *et al.* , 1988),(Lang *et al.* , 1995),(Olesen & Petersen, 1996)). For every 'unacceptable' projection of an inefficient DMU on a non-fully-dimensional facet these approaches try to identify the closest 'acceptable' facet and use that to specify a new target for the specific DMU. This new target is not a part of the original production possibility set but of an enlarged one that was created by extrapolating existing facets. In our case, including all reference points for cross-evaluations in the production possibility set would lead to a highly unrealistic enlargement. For example, in figure (1) even when we only try to include reference points relevant for DMU E, this results in a new production possibility set bounded from the left by hyperplane $f_5$ alone. If we wanted to include all reference points then it would not be possible to describe the new production possibility set with the existing hyperplanes.

Having established that cross-evaluation methodology is based on a significant departure from fundamental assumptions of DEA, we now attempt to restore its credentials by reflecting further on what is really at stake. Cross-evaluation was never intended for setting targets for DMUs, but rather with measuring efficiency with the use of scoring 'recipes' used by peer DMUs. Therefore, reference points need not necessarily be interpreted as targets and hence included in the production possibility set. We have shown before that the efficient frontier can be identified by a collection of hyperplanes that are defined by different sets of weights. Each hyperplane, when taken alone, corresponds to an alternative frontier that can be used to evaluate efficiencies. Consequently, instead of measuring efficiency by establishing a unique frontier, cross-evaluation measures efficiency relative to different frontiers and reports more than one efficiency scores, subsequently combined into a unique score for each DMU. The information included in the final score depends on the available information on individual weighting schemes.

There does remain, however, a fundamental weakness in cross evaluation methodology. Traditional cross-evaluation approaches can only identify and include a single weighting scheme for each DMU and thus are excluding a significant amount of information from the analysis, leading to inconsistent and

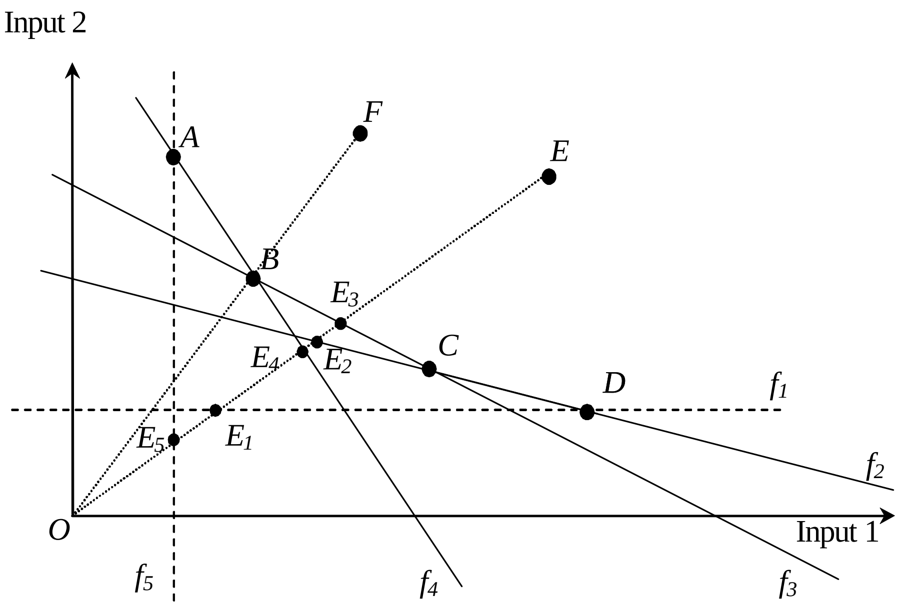


Figure 1: Cross-Evaluation and the DEA production possibility set

misleading interpretations. We illustrate these anomalies and present an alternative approach in later sections .

We now look at the VRS DEA model (Banker *et al.* , 1984). In this case things are more problematic. Consider the example in Figure 2. The DEA efficient frontier is given by the piecewise linear segment ABCD adjoined by the rays AA′ and DD′. Points $E_1$ and $E_2$ correspond to input oriented cross-evaluations of E while $E_3$ and $E_4$ correspond to output oriented cross evaluations (not all cross-evaluations are plotted). As in the CRS case, $E_2$ and $E_4$ lie outside of the DEA production possibility set, in fact they violate the convexity constraint of the VRS model. The additional problem concerns input oriented cross-evaluations on some hyperplanes[1]. For example, point $E_2$ has negative input data which implies a negative value for the particular cross-efficiency score. Another problem is that by allowing all cross-evaluations for a DMU we are evaluating it by using weights that are not always compatible with the returns-to-scale type exhibited by that DMU. This seems to contradict the motivation behind the VRS model in which DMUs that exhibit increasing returns-to-scale are never compared to

[1]In this model each hyperplane is of the form $u\mathbf{y}_j - v\mathbf{x}_j = \beta$, where $(u, v)$ is the normal to the hyperplane and $\beta$ is the offset term. The additional problem concerns some input-oriented cross-evaluations on decreasing returns-to-scale hyperplanes, i.e. the ones with a negative offset.

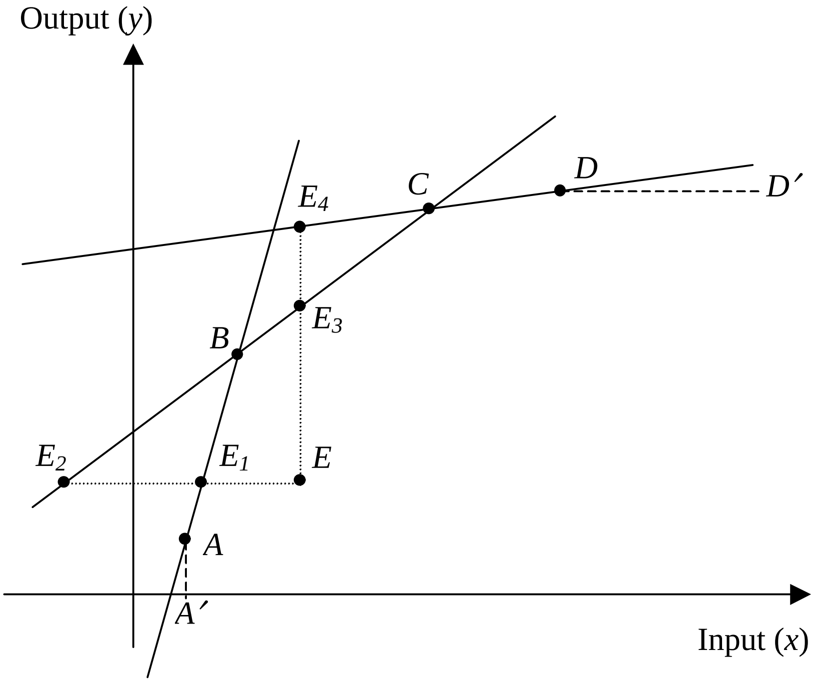


Figure 2: Cross-Evaluations and the VRS production set

DMUs that exhibit decreasing returns-to-scale and vice-versa (Tone, 1996). These two problems reduce the usefulness of cross-evaluation in VRS cases. Possible remedies could be to perform cross-evaluation with different (non-radial and non-oriented) efficiency measures and to calculate the cross-efficiencies of a DMU only with respect to DMUs that exhibit compatible types of returns-to-scale. These are left as open questions for further research and for the purposes of this article we confine ourselves to the CRS case.

# 4 A critical appraisal of traditional Cross-Evaluation Approaches

## 4.1 Implications of multiple optimal weighting schemes

It is very often the case that the solution of a DEA model is not unique, i.e. there exist multiple sets of weighting schemes that yield the same efficiency score for a DMU. The existence of multiple optimal weights has serious implications for cross-evaluation, the reason being that cross-efficiency scores are entirely dependent on which of the multiple weighting schemes is chosen. In addition, because of the nature of algorithms used to solve DEA models, in such cases only one of the multiple sets of weights can be discovered. Thus, although the use of different weighting schemes will produce different cross-efficiency scores, there is not even a way of being able to choose among some of the multiple optima. Traditional cross-evaluation approaches tackle this problem by introducing a second criterion, in the form of an objective function, in order to identify a unique set of weights for every DMU that either minimises or maximises the average cross-efficiency of all other DMUs. Depending on the criterion, we distinguish between aggressive and benevolent formulations. This section highlights some inadequacies of these traditional approaches. We start by identifying problems that are specific to the models and finish with some general remarks on the philosophy of these approaches.

## 4.2 Existing models

We will examine the aggressive formulations of four cross-evaluation approaches, taken from (Doyle & Green, 1995). All four approaches are implemented by two-stage models. In the first stage simple efficiency is maximised and in the second stage a secondary goal is introduced, which is to minimise the cross-efficiencies of all other DMUs in some way, as summarised in (8).

$$\text{Primary goal} \;\; \max \;\; h_k \tag{8a}$$

$$\text{Secondary goal} \;\; \min \;\; \frac{1}{n-1} \sum_{j \neq k} \frac{O_{jk}}{I_{jk}} \tag{8b}$$

In the above approach the secondary goal involves the minimisation of the average appraisal of peers by DMU $k$. Even if we ignore the fact that this goal is non-linear, there are problems concerning the choice among multiple optimal weights. Although this two stage approach has been suggested as a remedy for multiple optima in the first stage, it may be the case that there are multiple optimal solutions for the second stage, in which case the problem of having to choose one among many remains. As mentioned earlier, such a choice has implications because each weighting scheme will produce different cross-efficiency scores and therefore the average cross-efficiency scores will differ. Thus, the choice of weights will have a knock-on effect on the ranking of DMUs.

The non-linear secondary goal makes model (8b) computationally intractable. To tackle this computational difficulty, models (9) and (10), developed by Sexton *et al* (1986) and Doyle and Green (1994) respectively, were suggested as surrogates for model (8).

$$\max \ h_k \tag{9a}$$

$$\min \sum_{j \neq k} (O_{jk} - I_{jk}) \tag{9b}$$

$$\max \ h_k \tag{10a}$$

$$\min \frac{\sum_{j \neq k} O_{jk}}{\sum_{j \neq k} I_{jk}} \tag{10b}$$

Doyle and Green also report that (10) is a better surrogate and that the secondary goal (10b) is equivalent to minimising the cross-efficiency of a composite DMU which includes the 'economy' of all DMUs except DMU $k$. We denote this composite DMU by $C^k$ and calculate its input-output bundle, $(\mathbf{x}^k, \mathbf{y}^k)$, as follows:

$$(\mathbf{x}^k, \mathbf{y}^k) = (\sum_{j \neq k} \mathbf{x}_j, \sum_{j \neq k} \mathbf{y}_j) \tag{11}$$

We now discuss two problematic cases for (10) wherein we illustrate how even when all optimal weights are available, the surrogate model provides the wrong answer. Similar examples can be constructed for (9). In the first case we will be concerned with the ten DMUs given in table 2 and in the second case with the eight DMUs in table 3. We take $k = 1$ and denote by $\mathrm{C}_1^1 = (9,\ 36,\ 29)$ and $\mathrm{C}_2^1 = (7,\ 28,\ 28)$, the composite DMUs for the first and the second case respectively.

By using Fourier-Motzkin (F-M) elimination to obtain the extreme rays of the homogeneous cone of input-output weights as proposed in (Appa & Williams, 2006), we deduce that in both cases DMU 1 is efficient when using either of the two sets of weights given in Table 4, where $A_t^1$ is the average cross-efficiency of all DMUs except DMU 1 at weight $t$ (for $t = 1,\ 2$).

DMU 1 is an 'aggressive' DMU that wants to choose among its optimal weights in a way that minimises its average appraisal of peers ($A_t^1$). In the first case, if information on all weight sets were available, DMU 1 would simply choose the set that minimises $A_t^1$ which would be weight set two. However, by implementing surrogate model (10), DMU 1 will choose the weight set that minimises the

| DMU | Output (O) | Input 1 ($I_1$) | Input 2 ($I_2$) |
|---|---|---|---|
| 1 | 1 | 2 | 2 |
| 2 | 1 | 1 | 4 |
| 3 | 1 | 4 | 1 |
| 4 | 1 | 3 | 2 |
| 5 | 1 | 4 | 6 |
| 6 | 1 | 5 | 2 |
| 7 | 1 | 7 | 10 |
| 8 | 1 | 4 | 3 |
| 9 | 1 | 4 | 9 |
| 10 | 1 | 4 | 2 |

Table 2: Data for case 1

| DMU | Output (O) | Input 1 ($I_1$) | Input 2 ($I_2$) |
|---|---|---|---|
| 1 | 1 | 2 | 2 |
| 2 | 1 | 1 | 4 |
| 3 | 1 | 4 | 1 |
| 4 | 1 | 3 | 2 |
| 5 | 1 | 4 | 6 |
| 6 | 1 | 5 | 2 |
| 7 | 1 | 7 | 10 |
| 8 | 1 | 4 | 3 |

Table 3: Data for case 2

cross-efficiency of composite DMU $C_1^1$, which is set one. Clearly this is the wrong choice.

Case two illustrates the possibility of multiple optimal weights in the second stage of the optimisation procedure. By implementing (10), DMU 1 would arbitrarily choose between the two weight sets since they both render the same efficiency score for composite DMU $C_2^1$ although the average appraisal of peers is clearly minimised at weight set two. We mentioned earlier that the choice among multiple optimal weights in (8b) can affect the ranking of DMUs. This example shows that this can be the case for surrogate models as well, although, the problem in this case is worse since, as illustrated in case one, the wrong choice can be made when selecting arbitrarily among multiple optimal weights for a DMU.

Next consider another model given by :

| Weight Set ($t$) for DMU 1 | Weights | | | Efficiencies | | | |
|---|---|---|---|---|---|---|---|
| | O | $I_1$ | $I_2$ | $C_1^1$ | $A_t^1$ (Case 1) | $C_2^1$ | $A_t^1$ (Case 2) |
| 1 | 6 | 1 | 2 | 0.47368 | 0.60116 | 0.45714 | 0.52708 |
| 2 | 6 | 2 | 1 | 0.48648 | 0.56596 | 0.45714 | 0.45485 |

Table 4: Weight Sets and Efficiencies

$$\max \ h_{kk} \tag{12a}$$

$$\min \frac{O_{jk}}{I_{jk}} \ \forall \ k \neq j \tag{12b}$$

This final model given in (Doyle & Green, 1995) does not necessarily force DMUs to appraise all of their peers using the same weighting scheme. Instead, all DMUs are allowed to choose different weighting schemes for different cross-evaluations, such that all cross-efficiencies are minimised. However, one can argue that by allowing the use of different weights in different cross-evaluations DMUs are not consistent in appraising their peers. This also seems to contradict a statement in an earlier article of the same authors, where they stress that in an aggressive world "It is not enough to talk yourself up; you must talk the others down too, but without being inconsistent" (Doyle & Green, 1994). Nevertheless, this approach does not face any of the problems that the previous approaches do. This seems to suggest that the greater number of weighting schemes is a desirable property for a cross-evaluation analysis provided that the weights will be consistently applied when calculating cross-efficiencies

### 4.3 General problems

There are also some general problems with the philosophy of traditional cross-evaluation approaches. A first point to be made is that aggressive formulations often result in the selection of maverick weighting schemes by DMUs where possible. Recall that maverick weights are especially used by DMUs that specialise in a subset of inputs and outputs and therefore it is very likely that such weights will provide the most conservative appraisal of peer DMUs. Consider the example in figure (1). DMU D is efficient at the weights associated with hyperplanes $f_1$ and $f_2$ and will choose the weight set that minimises the appraisal of its peers which is the set associated with $f_1$. However, $f_1$ is not a fully-dimensional hyperplane but corresponds to a maverick weighting scheme which assigns a zero weight to input 1. The same observation holds for DMU A which can choose between $f_4$ and $f_5$. This behaviour of certain DMUs contradicts the motivation of cross-evaluation for reducing the effect of unrealistic weighting schemes. One could try to resolve this issue by introducing weight restrictions in the process of calculating average cross-efficiencies, i.e. not include cross-evaluations of DMUs at maverick weights in the averaging. This however does not come without further complications, as shown below.

In general, the limited available information on weighting schemes reduces usefulness of weight restrictions within the traditional cross-evaluation framework because we still end up with unrealistic scores for some DMUs. For example, it could be the case that some DMUs that are in reality efficient appear to be inefficient in all cross evaluations. Consider again the example in figure (1). The standard DEA procedure might end up having identified hyperplanes $f_1$, $f_3$ and $f_5$. If cross-evaluations on $f_1$ and $f_5$ are excluded from the cross-efficiency matrix, DMUs A and D are only scored against $f_3$ and are therefore found to be inefficient even though there exist possibly realistic weighting schemes for which we have no information ($f_2$ and $f_4$) that declare these DMUs efficient.

# 5 A New Approach for Cross-Evaluation

The previous section has presented a collection of problematic examples for traditional cross-evaluation approaches. In all these, we have assumed that all possible weighting schemes were available. This allowed us to check the validity of the results arrived at by the traditional approaches and, more importantly, to identify correct or more appropriate solutions for some cases. It is therefore clear that the availability of all weighting schemes would greatly enhance cross-evaluation analyses. We investigate this here.

## 5.1 Including all weighting schemes

We have demonstrated how the information on all sets of weights can provide a solution for (8) and thus eliminate the need for surrogate models. Motivated by the usefulness of complete frontier information we propose a new approach for cross-evaluation based on the availability of this information. Traditional cross-evaluation approaches are handicapped, in the sense that they assume the unavailability of all weighting schemes. Instead, they introduce a secondary criterion in order to choose one weighting scheme for each DMU. However, it could be the case that neither aggressive nor benevolent formulations are suitable options, or even that the choice between these might not be easy. Green *et al* (1996) report that in the case of ranking R&D projects, arguments can be made for the use of both options. Additionally, although aggressive or benevolent formulations seem to provide some orientation to the analysis, this disguises the fact that a choice between these has to be made if an analysis is to be carried out at all.

In particular, we have used the Fourier-Motzkin Method to obtain all hyperplanes that define the CRS production possibility set, as described in (Appa & Williams, 2006). The availability of PORTA[2] a public domain software based on the Fourier-Motzkin method makes it very easy to compute the frontier for practical problems. A number of other approaches exist (see e.g. (Raty, 2002),(Olesen & Petersen, 2003),(Briec & Leleu, 2003)) and, in theory, any of these can also be applied.

The standard definition of cross-efficiency (the efficiency of DMU $j$ when using the weights selected by DMU $k$) is not sufficient when all weighting schemes are available, simply because many DMUs are efficient at more than one set of weights. In addition, some weights might be optimal for more than one DMUs, which would result in many repeated entries in the cross-efficiency matrix. To overcome these problems we build on Appa and Williams' approach that evaluates DMUs across all weighting schemes as part of a new framework for the solution of DEA models, and extend cross-evaluation to include all these possible scoring combinations.

Formally, let $I_C = \{1, ..., t\}$ be an index set for the $t$ hyperplanes that explicitly characterise $\hat{T}_C$. The set of all possible pairs of input and output weights (weighting schemes) is:

$$F = \{(u_i, v_i)|\ \forall i \in I_C\ : (u_i, -v_i) \in P_C\} \tag{13}$$

Including all weights in the analysis gives rise to the following definition:

**Definition 2** *The cross-efficiency of DMU $j \in \mathcal{J}$ relative to weighting scheme $(u_i, v_i) \in F$, denoted $c_{ji}$ is given by:*

$$c_{ji} = \frac{u_i \mathbf{y}_j}{v_i \mathbf{x}_j} \tag{14}$$

Calculating $c_{ji}$ for all combinations of $i$ and $j$ results in a restructured, $n \times t$ - dimensional, cross-efficiency matrix like the one given in table 5. Standard cross-evaluation procedures only obtain information on a small subset[3] of weighting schemes and thus can only construct an incomplete cross-efficiency matrix. Our cross-efficiency matrix contains all possible scoring combinations. We can formally define

[2] code by Thomas Christof, Heidelberg University and Andreas Loebel, Konrad-Zuse-Zentrum fur Informatik (ZIB)

[3] In fact, as the problem size grows in dimensions the subset of weighting schemes obtained by standard cross-evaluation becomes increasingly smaller compared to the complete set of weighting schemes.

| DMUs ($j$) | Weight Set ($i$) | | | | | $\bar{c}_j$ |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | ... | $t$ | |
| 1 | $c_{11}$ | $c_{12}$ | $c_{13}$ | ... | $c_{1t}$ | $\bar{c}_1$ |
| 2 | $c_{21}$ | $c_{22}$ | $c_{23}$ | ... | $c_{2t}$ | $\bar{c}_2$ |
| 3 | $c_{31}$ | $c_{32}$ | $c_{33}$ | ... | $c_{3t}$ | $\bar{c}_3$ |
| ... | ... | ... | ... | ... | ... | ... |
| $n$ | $c_{n1}$ | $c_{n2}$ | $c_{n3}$ | ... | $c_{nt}$ | $\bar{c}_n$ |
| $f_i$ | $f_1$ | $f_2$ | $f_3$ | ... | $f_t$ | |

Table 5: The complete Cross-Efficiency Matrix

the average cross efficiency and the maverick index for DMU $j$ as follows:

$$\bar{c}_j = \frac{1}{t}\sum_i c_{ji} \tag{15}$$

$$m_j = \frac{h_j - \bar{c}_j}{\bar{c}_j} \tag{16}$$

This new approach is structurally very similar to previous approaches but fundamentally different in its philosophy. Instead of averaging the efficiency appraisals by all peers we are now looking at the average *efficiency over all possible weights* that peers could have used. We have demonstrated that analyses that disregard some weighting schemes are intrinsically incomplete and unreliable. Including all weighting schemes in a cross evaluation analysis has the following advantages over traditional cross-evaluation approaches:

(1) Because cross-evaluation almost always achieves a unique ranking of DMUs through the averaging of their cross-efficiency scores, the amount of information on weights included in the averaging is of utmost importance. By including all weight information in the analysis we are no longer restricted to establishing a unique weighting scheme for every DMU. Hence, we eliminate the need for aggressive or benevolent formulations and overcome *all* the problematic cases discussed earlier. We can therefore produce more meaningful average cross-efficiency scores, ranking of DMUs and maverick indices.

(2) We are now able to introduce weight restrictions by a priori identification of unrealistic weighting schemes which can be excluded from the analysis. Since all weighting schemes are available we need not worry about unrealistically evaluating some DMUs because of limited information. Hence, we can now practically eliminate the effect of unrealistic weighting schemes on the average

cross-efficiency scores.

(3) Unlike all the models presented earlier, we are no longer restricted to input and output oriented efficiency measures. With all the hyperplanes of the frontier available it is easy to perform cross-evaluations with a variety of different efficiency measures for all DMUs. Briec and Leleu (2003) consider the problem of an arbitrary norm projection on the efficient frontier. They introduce the concept of the Hölder distance function but also employ the directional distance function ((Chambers *et al.* , 1996),(Chambers *et al.* , 1998)), and provide the framework for calculating these when all frontier hyperplanes have been identified. Such efficiency measures could be of particular value in overcoming the serious technical drawbacks outlined in section (3) for cross-evaluation under the VRS model.

### 5.2 Identifying maverick DMUs

We now describe how the new approach can help identify maverick DMUs more effectively. The maverick index in (16) compares a DMU's simple efficiency with its average cross-efficiency. The reasoning behind this is that if DMUs are using unrealistic weighting schemes the difference between these two efficiency measures will be high. This however does not take the behaviour of other DMUs into account. We maintain that it would be unreasonable to establish a DMU's behaviour as unrealistic without also considering how its peers behave. Consider the 2-input, 1-output CRS example in figure (3) where seventeen DMUs are plotted. By using $m_j$ in (16) we obtain a relatively high maverick index for DMUs A and D, but in doing so we have failed to recognise that there is a large number of DMUs (those in the cone spanned by OC and OD) that behave in a similar way to DMU D, i.e. they would all choose one of D's optimal weighing schemes. With this in mind, it would be dubious to designate D as a maverick DMU or the specific weighting scheme as unrealistic. We propose a new indicator that takes the behaviour of peer DMUs into account.

Let the set of optimal weighting schemes for DMU $j$ be:

$$F_j = \{(u_i, v_i) \in F |\; c_{ji} = \max_i \{c_{ji}\}\} \tag{17}$$

Let the average efficiency on weighting scheme $i$ be:

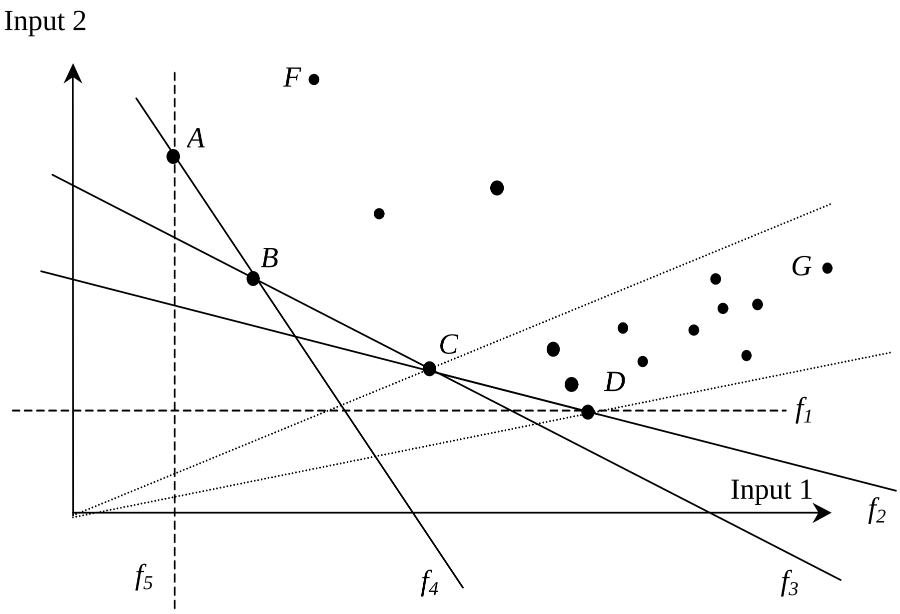


Figure 3: Identification of mavericks and under-achievers

$$\bar{f}_i = \frac{1}{n}\sum_j c_{ji} \tag{18}$$

Now for every DMU $j$ we define:

$$\kappa_j = \max\{\bar{f}_i | (u_i, v_i) \in F_j\} \tag{19}$$

Finally we define the new indicator as:

$$p_j = \frac{h_j - \kappa_j}{\kappa_j} \tag{20}$$

In trying to consider the behaviour of all DMUs we have based the new indicator on $\kappa_j$. The rationale for this is that if a relatively high number of DMUs use the same weighting scheme as DMU $j$, $\kappa_j$ would have a high value and therefore that DMU $j$ is not a maverick. Obviously a low $\kappa_j$ value relative to $h_j$ will lead to a high value for $p_j$, pointing to maverick DMUs. For example in figure (3) there is a high number of DMUs that achieve their maximum efficiency at weighting scheme $f_2$ and hence we could not identify DMUs C and D that choose $f_2$, as mavericks. DMU A on the other hand chooses weighting schemes that provide particularly low appraisals. Therefore the value of $p_j$ would be lower for DMUs C and D than for DMU A. As a note, we would like to say that there are alternative ways to define $\kappa_j$. Instead of taking the maximum appraisal by DMU $j$'s optimal weighting schemes one could take the minimum or even the average appraisal and define $p_j$ appropriately.

Both the conventional maverick index and the new indicator try to identify unrealistic behaviour but they do so in fundamentally different ways. Essentially, $m_j$ is examining the efficiency scores of a specific DMU on all weighting schemes and $p_j$ is examining the efficiency scores of all DMUs on a specific weighting scheme. We could also use these indices in parallel in order to better identify maverick DMUs.

### 5.3 Identifying under-achieving DMUs

Another feature of the new approach is its ability to identify under-achieving DMUs. It is important to differentiate between a low-achiever and an under-achiever among DMUs. The former is simply a DMU with a low simple efficiency score whereas the latter is a DMU with simple efficiency considerably lower than that of many other DMUs evaluated at its optimal weighting schemes. Under-achievement is not a characteristic of only very inefficient DMUs in the same way as maverickness is not a characteristic of only efficient DMUs. In figure (3) both DMUs F and G are low achievers but unlike F, G is also an under-achiever. Under-achieving DMUs can be identified by their $p_j$ values. Notice that this indicator can also take negative values (when $h_j < \kappa_j$) which is the case if the weighting scheme associated with $\kappa_j$, chosen by DMU $j$, provides higher appraisals to other DMUs than DMU $j$ on average. A significantly low $p_j$ value translates to many DMUs achieving considerably higher efficiency scores than DMU $j$ at a weighting scheme of its choice, so that DMU $j$ can be identified as an under-achiever.

### 5.4 Identifying maverick weighting schemes

The motivation behind our approach is also relevant when trying to identify maverick weighting schemes and impose weight restrictions. There is a considerable amount of literature on weight restrictions in DEA and a variety of methods for imposing these[4]. The focus in all these is in examining the values of the input and output multipliers in each weighting scheme. The decision-maker usually has a view

[4] See Thanassoulis *et al* (2004) for a review of methods for imposing weight restrictions in DEA.

of how the production process taking place in DMUs should be represented in the values of the weights selected by DMUs. If, for a specific weighting scheme, these values are not in line with the views held by the decision-maker then that weighting scheme is termed unrealistic. Clearly, obtaining weighting schemes with realistic values for input and output weights is extremely important. However, in figure (3) we illustrated why weight restriction analysis should be complemented with information about how DMUs behave with respect to different weighting schemes, an issue that has been relatively unexplored. We suggest that one way to do this is by considering the average efficiency for each weighting scheme. Including such information can prove very useful in identifying unrealistic weighting schemes. For example, in cases where merely considering the weight values cannot establish whether a weighting scheme is unrealistic, we can get a better insight by considering how favourable this scheme is to DMUs. Additionally, identifying weighting schemes that provide relatively high appraisals can give us an insight on the 'acceptable' range of values for the input and output weights. For a central organisation attempting to improve performance of branches the acceptability of weights used in making comparisons is of critical importance. The $\bar{f}_i$ measure has the potential as a tool for selecting a restricted set of weights in an objective and acceptable manner. We leave this idea open for further research.

### 5.5 Illustrative example

We illustrate all the above points with a simple 2-input, 1-output example using the hypothetical data for fifteen DMUs given in table (6). The production possibility set for this small example is given in figure (4). It is perhaps easily observable that most DMUs make more efficient use of input 1 than input 2. The efficient frontier is comprised of six facets of which the corresponding weighting schemes are given in table (7)[5].

Having identified all weighting schemes we can now consider whether some of them are unrealistic. Two obvious candidates are weighting schemes 1 and 6 which include zero multipliers and hence correspond to non fully-dimensional weighting schemes. Other possible candidates include weighting schemes 2 and 5 in which the weights for input 2 and input 1 respectively are very small. As suggested before we will also consider the average efficiency appraisals of these weighting schemes (column $\bar{f}_i$ in table 9). Weighting scheme 2 has one of the highest average appraisals and weighting scheme 5 one of the lowest.

[5] Note that these can be scaled by any positive multiplier

given the behaviour of DMUs in our dataset, namely the fact that most of them make more efficient use of input 1, this comes as no surprise. Therefore identifying weighting scheme 2 as unrealistic would be dubious but on the other hand we now have more evidence to suggest that weighting scheme 5 is unrealistic. Once we have established which weighting schemes will be excluded we can proceed with the calculation of the cross-efficiency matrix. The complete cross-efficiency matrix for this data-set is given in the appendix. Here we will examine three cases. In case one we have included all weighting schemes, in case two we have excluded weighting schemes 1 and 6 and in case three we have additionally excluded weighting scheme 5. In table 8 we present the results on simple and average cross-efficiencies, rankings, and the $m_j$ and $p_j$ values for all DMUs in the three cases.

It is of particular interest to compare the results for $m_j$ and $p_j$. In case one DMUs 1 and 7 achieve their simple efficiency score by using non fully-dimensional weighting schemes, so that it is no surprise that they achieve the two highest $m_j$ values with $m_1$ slightly higher than $m_7$.This fails to grasp that there are many other DMUs behaving similarly to DMU 1, in the sense that they achieve their simple efficiency by assigning greater importance to input 1, but very few DMUs behave similarly to DMU 7. This is reflected in the values of $p_j$ for DMUs 1 and 7 (calculated with use of weighting schemes 1 and 6 respectively) which still achieve the highest $p_j$ values but with the difference that $p_7$ is much higher than $p_1$, i.e. DMU 7 is more of a maverick than DMU 1. The same observation holds for DMUs 2 and 6 which achieve very similar $m_j$ values but very different $p_j$ values with $p_6$ being much higher than $p_2$ ($p_2$ is calculated with weighting scheme 2 and $p_6$ with weighting scheme 5).

In case two DMUs 1 and 7 are still the most maverick DMUs in the context of index $m_j$ but for $p_j$ this is no longer the case. This is due to the different ways in which the exclusion of weighting schemes affects the two indices in general. The values of $m_j$ change for all DMUs since this index is affected by all available weighting schemes. On the other hand, for any DMU, $p_j$ is only affected by one weighting scheme which explains why only $p_1$ and $p_7$ decrease (calculated with weighting schemes 2 and 5 respectively) and all other indices are unaffected. In the context of $p_j$ these two DMUs are now using more realistic weighting schemes for their self appraisal so they now possess less maverick characteristics. What is more, DMU 1 which has a big family of peers with similar behaviour achieves a lower $p_j$ value than DMU 7 and DMU 6, two of the relatively few DMUs which make more efficient use of input 2. This also explains why DMU 5 achieves a relatively low value in index $m_j$ but not in $p_j$.

Finally, in case three the exclusion of weighting scheme 2 forces DMUs 6 and 7 to achieve their

| DMU | Input 1 ($I_1$) | Input 2 ($I_2$) | Output (O) |
|---|---|---|---|
| 1 | 3 | 17 | 1 |
| 2 | 3 | 15 | 1 |
| 3 | 4 | 11 | 1 |
| 4 | 6 | 6 | 1 |
| 5 | 10 | 4 | 1 |
| 6 | 14 | 3 | 1 |
| 7 | 16 | 3 | 1 |
| 8 | 4 | 16 | 1 |
| 9 | 5 | 15 | 1 |
| 10 | 7 | 15 | 1 |
| 11 | 5 | 17 | 1 |
| 12 | 6 | 17 | 1 |
| 13 | 6 | 19 | 1 |
| 14 | 8 | 19 | 1 |
| 15 | 8 | 11 | 1 |

Table 6: Data for fifteen DMUs

| Weighting Scheme ($i$) | Weights | | | Efficient DMUs | $\bar{f}_i$ |
|---|---|---|---|---|---|
| | $I_1$ | $I_2$ | O | | |
| $f_1$ | 0.3333 | 0 | 1 | 1, 2 | 0.53869 |
| $f_2$ | 0.1481 | 0.0370 | 1 | 2, 3 | 0.71478 |
| $f_3$ | 0.1190 | 0.0476 | 1 | 3, 4 | 0.73164 |
| $f_4$ | 0.0556 | 0.1111 | 1 | 4, 5 | 0.61701 |
| $f_5$ | 0.0385 | 0.1538 | 1 | 5, 6 | 0.54137 |
| $f_6$ | 0 | 0.3333 | 1 | 6, 7 | 0.36188 |

Table 7: Weighting Schemes for DMUs in table (6)

simple efficiency with use of weighting scheme 4. This means that their behaviour is moving closer to the large family of DMUs that assign greater weights to input 1, so that their $p_j$ values decrease with $p_6$ once again lower than $p_1$.

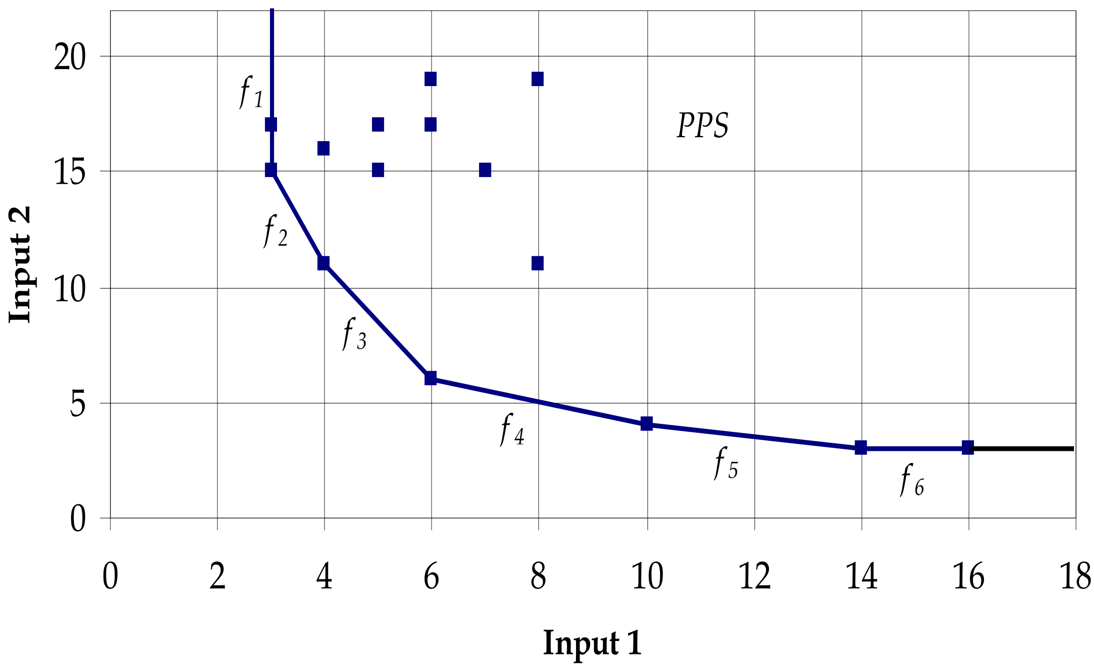


Figure 4: The production possibility set for DMUs in table (6)

| | Case 1 | | | | | Case 2 | | | | | Case 3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DMU | $h_j$ | $\bar{c}_j$ | rank | $m_j$ | $p_j$ | $h_j$ | $\bar{c}_j$ | rank | $m_j$ | $p_j$ | $h_j$ | $\bar{c}_j$ | rank | $m_j$ | $p_j$ |
| 1 | 1 | 0.6362 | 7 | 0.5778 | 0.8563 | 0.931 | 0.6602 | 6 | 0.4102 | 0.7283 | 0.931 | 0.7582 | 6 | 0.2279 | 0.7283 |
| 2 | 1 | 0.6819 | 5 | 0.4665 | 0.3990 | 1 | 0.7229 | 5 | 0.3834 | 0.3990 | 1 | 0.8263 | 5 | 0.2103 | 0.3990 |
| 3 | 1 | 0.7095 | 3 | 0.4095 | 0.3668 | 1 | 0.8085 | 3 | 0.2369 | 0.3668 | 1 | 0.7984 | 3 | 0.1143 | 0.3668 |
| 4 | 1 | 0.7945 | 1 | 0.2587 | 0.3668 | 1 | 0.9417 | 1 | 0.0619 | 0.3668 | 1 | 0.9667 | 1 | 0.0345 | 0.3668 |
| 5 | 1 | 0.7313 | 2 | 0.3675 | 0.6207 | 1 | 0.8344 | 2 | 0.1984 | 0.6207 | 1 | 0.7792 | 2 | 0.2833 | 0.6207 |
| 6 | 1 | 0.6874 | 4 | 0.4547 | 0.8472 | 1 | 0.7276 | 4 | 0.3745 | 0.8472 | 0.9 | 0.6467 | 4 | 0.4135 | 0.6624 |
| 7 | 1 | 0.6376 | 6 | 0.5683 | 1.7633 | 0.9286 | 0.6596 | 7 | 0.4079 | 1.5660 | 0.8182 | 0.5699 | 7 | 0.4358 | 1.2610 |
| 8 | 0.8438 | 0.5786 | 8 | 0.4584 | 0.1805 | 0.8438 | 0.6335 | 8 | 0.3320 | 0.1805 | 0.8438 | 0.7172 | 8 | 0.1766 | 0.1805 |
| 9 | 0.7714 | 0.5416 | 9 | 0.4224 | 0.0792 | 0.7714 | 0.6123 | 9 | 0.2598 | 0.0792 | 0.7714 | 0.6831 | 9 | 0.1293 | 0.0792 |
| 10 | 0.6462 | 0.4629 | 13 | 0.3690 | -0.1168 | 0.6462 | 0.5372 | 11 | 0.2030 | -0.1168 | 0.6462 | 0.5869 | 12 | 0.1011 | -0.1168 |
| 11 | 0.7297 | 0.5060 | 11 | 0.4422 | 0.0209 | 0.7297 | 0.5648 | 12 | 0.2919 | 0.0209 | 0.7297 | 0.6344 | 11 | 0.1503 | 0.0209 |
| 12 | 0.6585 | 0.4654 | 12 | 0.4148 | -0.0787 | 0.6585 | 0.5290 | 13 | 0.2447 | -0.0787 | 0.6585 | 0.5882 | 13 | 0.1195 | -0.0787 |
| 13 | 0.6279 | 0.4383 | 14 | 0.4327 | -0.1215 | 0.6279 | 0.4929 | 14 | 0.2738 | -0.1215 | 0.6279 | 0.5515 | 14 | 0.1385 | -0.1215 |
| 14 | 0.5385 | 0.3836 | 15 | 0.4038 | -0.2640 | 0.5385 | 0.4422 | 15 | 0.2178 | -0.2640 | 0.5385 | 0.4864 | 15 | 0.1071 | -0.0741 |
| 15 | 0.6774 | 0.5088 | 10 | 0.3313 | -0.0741 | 0.6774 | 0.6013 | 10 | 0.1265 | -0.0741 | 0.6774 | 0.6351 | 10 | 0.0666 | -0.0741 |

Table 8: Results

### 5.6 Computational considerations

Before concluding this article it would be appropriate to comment on the computational difficulty of this approach. The problem of identifying all weighting schemes is equivalent to explicitly characterising the DEA polyhedral production possibility set with a collection of hyperplanes. There are a number of available algorithms that can do that but they all suffer from exponential growth as the dimension of the problem (inputs, outputs and DMUs) grows. There is, however, a two-phase approach (see (Olesen & Petersen, 2003) and (Appa & Williams, 2006)) that manages the problem of exponential growth well. The essential idea is to solve at most one linear-programme for each DMU in phase one to identify the set $\mathcal{E}$ of all efficient DMUs. Since only efficient DMUs can span the rays of the convex cone of optimal weights (or equivalently span facets of the DEA efficient frontier), in phase two the computationally expensive algorithm for finding the convex cone is run on data for DMUs in set $\mathcal{E}$ only. In practice we have found that for problems with $|\mathcal{E}| \leq 50$ and $s + m \leq 10$ PORTA was able to provide the complete set of optimal weights in reasonable time. As part of a project to evaluate the efficiency of schools in England we were able to solve this problem for more than 2500 schools using a model with six inputs and two outputs. It would be an interesting direction for future research to identify algorithms and approaches that can deliver the set of all weighting schemes fro larger problems take advantage of the special characteristics of the DEA problems.

## 6 Conclusion

One can argue that cross-evaluation is a cross-breed between standard DEA and efficiency evaluation by externally imposed criteria, and that by combining the two it inherits desirable attributes from both. On the one hand, by not allowing total weight-flexibility, it tackles the problem of high simple efficiency scores based on unrealistic weighting schemes. However, the weights are not arbitrarily invented and imposed by some external agency but have been established through a detailed analysis of the dataset, and as in the standard DEA, have been generated by the dataset. In this article we have identified some flaws in the traditional approaches to cross-evaluation and suggested, at least for the CRS model, a new approach based on computing and using the complete set of weights. The task of extending our work to take account of returns to scale and of using cross-evaluation for non-radial models remains undone. We hope, however, that we have provided the necessary tools for a growth in applications of the technique in the real world.

# Appendix

| W. Scheme | DMUs | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| $f_1$ | 1 | 1 | 0.75 | 0.5 | 0.3 | 0.2143 | 0.1875 | 0.75 | 0.6 | 0.4286 | 0.6 | 0.5 | 0.5 | 0.375 | 0.375 |
| $f_2$ | 0.931 | 1 | 1 | 0.9 | 0.6136 | 0.4576 | 0.403 | 0.8438 | 0.7714 | 0.6279 | 0.7297 | 0.6585 | 0.6279 | 0.5294 | 0.6279 |
| $f_3$ | 0.8571 | 0.9333 | 1 | 1 | 0.7241 | 0.5526 | 0.4884 | 0.8077 | 0.7636 | 0.6462 | 0.7119 | 0.6562 | 0.6176 | 0.5385 | 0.6774 |
| $f_4$ | 0.4865 | 0.5455 | 0.6923 | 1 | 1 | 0.9 | 0.8182 | 0.5 | 0.5143 | 0.4865 | 0.4615 | 0.45 | 0.4091 | 0.3913 | 0.6 |
| $f_5$ | 0.3662 | 0.4127 | 0.5417 | 0.8667 | 1 | 1 | 0.9286 | 0.3824 | 0.4 | 0.3881 | 0.3562 | 0.3514 | 0.3171 | 0.3095 | 0.5 |
| $f_6$ | 0.1765 | 0.2 | 0.2727 | 0.5 | 0.75 | 1 | 1 | 0.1875 | 0.2 | 0.2 | 0.1765 | 0.1765 | 0.1579 | 0.1579 | 0.2727 |

Table 9: Cross-Efficiency Table for DMUs in table (6)

## References


Adler, N., Friedman, L., & Sinuany-Stern, Z. 2002. Review of Ranking Methods in the Data Envelopment Analysis Context. *European Journal of Operational Research*, **140**, 249–265.

Ali, A.I., & Seiford, L.M. 1993. The Mathematical Programming Approach to Efficiency Analysis. *Chap. 3, pages 120–159 of:* Fried, H.O., Lovell, C.A.K., & Schmidt, S.S. (eds), *The Measurement of Productive Efficiency*. Oxford University Press.

Anderson, T.R., Hollingsworth, K., & Inman, L. 2002. The Fixed Weighting Nature of A Cross-Evaluation Model. *Journal of Productivity Analysis*, **17**, 249–255.

Angulo-Meza, L., & Lins, M.P.E. 2002. Review of Methods for Increasing Discrimination in Data Envelopment Analysis. *Annals of Operations Research*, **116**, 225–242.

Appa, G, & Williams, H.P. 2006. A New Framework for the Solution of DEA Models. *European Journal of Operational Research*, **172**(2), 604–615.

Banker, R.D., Charnes, A., & Cooper, W.W. 1984. Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science*, **30**(9), 1078–1092.

Bessent, A., Bessent, W., & Elam, J. 1988. Efficiency Frontier Determination by Constrained Facet Analysis. *Operations Research*, **36**(5), 785–796.

Briec, W., & Leleu, H. 2003. Dual Representations of Non-Parametric Technologies and Measurement of Technical Efficiency. *Journal of Productivity Analysis*, **20**, 71–96.

Chambers, R., Chung, Y., & Fare, R. 1996. Benefit and Distance Functions. *Journal of Economic Theory*, **70**, 407–419.

Chambers, R., Chung, Y., & Fare, R. 1998. Profit, Directional Distance Functions and Nerlovian Efficiency. *Journal of Optimization Theory and Applications*, **98**(2), 351–364.

Charnes, A., Cooper, W.W., & Rhodes, E. 1978. Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research*, **2**, 429–444.

Charnes, A., Cooper, W.W., Huang, Z.M., & Sun, D.B. 1990. Polyhedral Cone-Ratio DEA Models with with an Illustrative Application to Large Commercial Banks. *Journal of Econometrics*, **46**, 73–91.

Chen, T. 2002. An Assesment of Technical Efficiency and Cross-Efficiency in Taiwan's Electricity Distribution Sector. *European Journal of Operational Research*, **137**, 421–433.

DfES. 2005. *Private Communication*.

Doyle, J., & Green, R. 1994. Efficiency and Cross-Efficiency in DEA; Derivations, Menings and Uses. *The Journal of The Operational Research Society*, **45**(5), 567–578.

Doyle, J.R. 1995. Multiattribute Choice for the Lazy Decision Maker: Let the Alternatives Decide. *Organizational Behavior and Human Decsision Processes*, **62**(1), 87–100.

Doyle, J.R., & Green, R.H. 1995. Cross-Evaluation in DEA: Improving Discrimination Among DMUs. *Information Systems and Operational Research*, **33**(3), 205–222.

Dyson, R.G., & Thanassoulis, E. 1988. Reducing Weight Flexibility in Data Envelopment Analysis. *Journal of the Operational Research Society*, **39**, 563–576.

Ertay, T., & Ruan, D. 2005. Data Envelopment Analysis Decision Model for Optimal Operator Allocation in CMS. *European Journal of Operational Research*, **164**, 800–810.

Farrell, M.J. 1957. The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society*, **120**(3), 253–281.

Green, R.H., & Doyle, J.R. 1995. On Maximising Discrimination in Multiple Criteria Decision Making. *The Journal of the Operational Research Society*, **46**(2), 192–204.

Green, R.H., Doyle, J.R., & Cook, W.D. 1996. Preference Voting and Project Ranking Using DEA and Cross-Evaluation. *European Journal of Operational Research*, **90**, 461–472.

Gregoriou, G.N., Sedzro, K., & J.Zhu. 2005. Hedge Fund Performance Appraisal Using Data Envelopment Analysis. *European Journal of Operational Research*, **164**, 555–571.

Lang, P., Yolalan, O.R., & O.Kettani. 1995. Controlled Envelopment by Face Extension in DEA. *Journal of the Operational Research Society*, **46**(4), 473–491.

Mavrotas, G., & Trifillis, P. 2006. Multicriteria Decision Analysis with Minimum Information: Combining DEA with MAVT. *Computers and Operations Research*, **33**, 2083–2098.

Olesen, O.B., & Petersen, N.C. 1996. Indicators of Ill-Conditioned Data Sets and Model Misspecification in Data Envelopment Analysis: An Exteded Facet Approach. *Management Science*, **42**(2), 205–219.

Olesen, O.B., & Petersen, N.C. 2003. Identification and Use of Efiicient Faces and Facets in DEA. *Journal of Productivity Analysis*, **20**, 323–360.

Oral, M., Kettani, O., & Lang, P. 1991. A Methodology for Collective Evaluation and Selection of Industrial RD Projects. *Mnagement Science*, **37**(7), 871–885.

Raty, T. 2002. Efficient Facet Efficiency Index: A Variable Returns to Scale Specification. *Journal of Productivity Analysis*, **17**, 65–82.

Sarkis, J. 2000. A Comparative Analysis of DEA as a Discrete Alternative Multiple Criteria Decision Tool. *European Jourmal of Operational Research*, **123**, 543–557.

Sarkis, J, & S.Talluri. 2004. Performance Based Clustering for Benchmarking of US Airports. *Transportation Research Part A-Policy and Practice*, **38**(5), 329–346.

Sexton, T.R., Silkman, R.H., & Hogan, A. 1986. Data Envelopment Analysis: Critique and Extensions. *Pages 73–105 of:* Silkman, R.H. (ed), *Measurin Efficiency: An Assessment of Data Envelopment Analysis.* San Francisco: Jossey Bass.

Shang, J., & Shueyoshi, T. 1995. A Unified Framework for the Selection of a Flexible Manufacturing System. *European Journal of Operational Research*, **85**, 297–315.

Sueyoshi, T. 1999. DEA Duality on Returns To Scale (RTS) in Production Cost Analyses: An Occurence of Multiple Solutions and Differences Between Production-Based and Cost-Based RTS Estimates. *Management Science*, **45**(11), 1593–1608.

Thanassoulis, E., Portella, M.C.P., & Allen, R. 2004. Incorporating Value Judgements in DEA. *Chap. 4, pages 99–138 of:* Cooper, W.W., Seiford, L. M., & Zhu, J. (eds), *Handbook on Data Envelopment Analysis.* Kluwer Academic Publishers.

Thompson, R.G., Jr, F.D. Singleton, Thrall, R.M., & Smith, B.A. 1986. Comparative Site Evaluations for Locating a High-Energy Physics Lab in Texas. *Interfaces*, **16**, 35–49.

Tone, K. 1996. A Simple Characterization of Returns To Scale in DEA. *Journal of the Operations Research Society of Japan*, **39**, 604–613.