Polyhedral attributes of Production Possibility Sets in Data Envelopment Analysis, with applications to Sensitivity Analysis and Cross-Evaluation Methodologies

Nikolaos Argyris

Operational Research Group, Department of Management

London School of Economics and Political Science

Thesis submitted to the Department of Management

of the London School of Economics, for the degree of Doctor of Philosophy

October 2007

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

Abstract

In this thesis we study some critical problems in the area of Data Envelopment Analysis (DEA) within the unifying framework of polyhedral characteristics of the production possibility sets and efficiency frontiers of important DEA models. Recent developments in DEA have made it possible to identify the efficient frontier explicitly. This thesis builds on these developments to make the following contributions.

We establish theoretical results on the efficiency classifications of surfaces of the boundaries of the production possibility sets. These systematise existing research in the field and fill in many gaps. Our main results provide necessary and sufficient conditions for characterising fully-dimensional efficient surfaces. In addition, the new theoretical framework leads us to discover and address inconsistencies in the related literature.

Next we study the sensitivity of efficiency classifications of Decision Making Units (DMUs) to data perturbations. In contrast to existing approaches, we study the effects of *arbitrary* data perturbations on the efficiency classifications of *all* DMUs. Theoretical constructs based on the polyhedral nature of production possibility sets lead to identifying a *Conditional Stability Region* for each DMU within which its data can be perturbed without affecting the efficiency classification of *any* other DMU.

Finally, we develop a new methodology for cross-evaluation in DEA which replaces the traditional approach of peer evaluation by evaluating DMUs across all possible weights obtained from our explicit identification of the DEA production possibility set. The new approach eliminates some major flaws and weaknesses of the traditional approach and produces more meaningful results. Moreover, a set of extensions to the new approach lead to tools that allow identification of DMUs with unrealistic efficiency scores as well as the identification of under-achieving DMUs, a

concept that is introduced here.

Acknowledgements

I would first like to express my gratitude to my supervisor Professor Gautam Appa. Over the past four years, his attentive attitude and academic rigour have had a major influence on my research and have inspired my own academic attitude. In addition to his academic support, I would like to thank him for the support that he has provided on a personal level.

I am also grateful to Professor Paul Williams and Dr Susan Powell for their helpful comments and advice on various academic matters during the past four years. As a supervisor of my Master's dissertation, Dr Susan Powell has played an important role in motivating me to begin my doctoral studies in DEA.

During my studies I have been lucky to have had contact with two talented DEA practitioners. Darren Pigg and Andy Webster, from the earstwhile Department for Education and Skills (DfES), have provided helpful practical insights to DEA-related issues and I have always been motivated by our discussions on these.

The Operational Research (OR) Group at the LSE is not only a very stimulating environment but also a pleasant place for conducting research. This is in no small part due to the two departmental administrators, Brenda Mowlam and Jenny Robinson. Over the past four years they have made my life easier not only by patiently helping with all sorts of administrative requests, but also by always being eager to listen to my concerns and provide useful advice.

I would also like to thank all of my fellow PhD students, especially Kai Becker, Joanna Katrantzi, Kostas Papalamprou and Srini Parthasarathy, for their friendship and support at various times during the past four years.

I am grateful to the DfES for funding my doctoral studies via a research project. Thanks are also due to the OR Group's various decision makers for putting me forward for financial support

from the LSE and also for giving me the opportunity to teach various classes, something that I have enjoyed immensely.

I want to express my deepest gratitude to my fiancée Lynne for the devoted attention and support that she has offered to me over these four years. Despite the peculiar schedule and nature of my work she has always shown patience and understanding. For this and so much more, I am indebted to her.

It is hard for me to find words to express my gratitude to my family: my parents and my younger brother. Throughout my life they have always offered me their support unconditionally, even when this required sacrifices on their part. I feel the greatest debt towards them; it is no exaggeration to say that I owe them everything.

Στούς γονείς μου και τον αδερφό μου.

Introduction

In this thesis we use a unified framework relating to polyhedral attributes of the production possibility sets constructed by DEA models, to study some important problems in DEA. The contributions that we make are twofold. First, we establish a thorough set of theoretical results that covers many 'gaps' in the field and provides new insights to these problems. Second, by utilising this theoretical framework we extend the traditional scope of the areas under study as well as introduce new tools and methodologies for these.

Before I describe the contents of this thesis in more detail, I would like to mention how this PhD topic came into existence. During my Master's dissertation, at the same department where I pursued my doctoral studies, I was undertaking a student project at the Department for Education and Skills (DfES). This examined the use of DEA for evaluating the efficiency of secondary Schools in England. It was during this time that I got properly acquainted with DEA and became very interested in the subject. At that time, my two PhD supervisors were introducing a new framework for the solution of DEA models that was based on the explicit identification of DEA production possibility sets. Following the successful completion of my project, an opportunity for a DfES funded PhD came up, under the supervision of Professor Gautam Appa. It was agreed that this would look at how the aforementioned framework can be used to extend existing methodologies in DEA as well as develop new ones, and the project was given the title 'Developing New Techniques for DEA'. This thesis describes the theory and methodology of the techniques that were developed during this project.

The first chapter of the thesis serves as an introduction to general economic efficiency concepts as well as DEA. In addition, we use this chapter to introduce the basic notation used in this thesis, as well as all the models and theory required for the understanding of the following chapters.

The DEA production possibility sets are polyhedra, and the efficient frontiers used in DEA are, broadly speaking, the boundaries of these polyhedra. Studying the polyhedral characteristics of DEA production possibility sets is a relatively recent trend in the DEA field. This is precisely what we undertake in chapters two and three, in great detail.

In chapter two we start by introducing the necessary theory for studying polyhedral aspects of production possibility sets. This includes some basic concepts from linear algebra and polyhedral theory. We then introduce two alternative representations of production possibility sets, viz. facets of the polyhedron and all extreme points and rays of the polyhedron. This is carried out for both Constant and Variable Returns-to-Scale models. Finally, we examine the problem of explicit identification of production possibility sets and review the existing approaches for this.

In chapter three we provide a thorough theoretical framework relating to the efficiency classifications and the dimensions of surfaces of production possibility sets. This systematises existing research in the field and fills in many gaps. Our main results provide necessary and sufficient conditions for characterising fully-dimensional efficient surfaces of the boundaries of production possibility sets. Using our new results we discover some inconsistencies in the existing literature on this topic. We provide a general description of these problems in the context of our new results, and counter-examples by way of rigorous proof.

Sensitivity analysis in DEA has always been a popular research topic. In chapter four we review the existing approaches for sensitivity analysis, examine their strengths and weaknesses, and point out that all of these focus on very restrictive cases of this problem, where it is assumed that only very specific types of data perturbations can occur.

This limited view of sensitivity analysis is addressed in chapter five. The problem that we study concerns the sensitivity analysis of efficiency classifications of DMUs by DEA, to data perturbations. We introduce a new approach, termed *Conditional Stability Analysis* (CSA), which is more powerful than existing approaches. It enables a study of the effect of *arbitrary* perturbations on the stability of efficiency classifications. In addition, CSA is not merely confined to considering the classification of only perturbed DMUs. It leads instead to the identification of geometrical regions, called conditional stability regions, within which a DMU's data can be perturbed without affecting the original efficiency classifications of *all* DMUs. We develop a set of theoretical constructs for this which are based on the polyhedral attributes of DEA production possibility sets. These results

are utilised to provide exhaustive characterisations of conditional stability regions. We also put forward a procedure for the computation of conditional stability regions. This is designed to reduce the problem of explicit identification of production possibility sets by focusing on local parts of these and has the potential to considerably reduce the overall computational effort.

In the final two chapters of this thesis we study the concept of cross-evaluation in DEA. In chapter six we conduct a critical investigation of its current theoretical framework and the computational tools used by traditional cross-evaluation approaches, which rely on restrictive formulations of this problem. Our analysis reveals some important flaws and weaknesses which we demonstrate by providing examples.

In chapter seven we point out that the general philosophy for existing cross-evaluation approaches is flawed. We address this by introducing a new methodology for cross-evaluation in DEA. The new approach departs from the traditional rationale of peer appraisal and focuses on the evaluation of DMUs across *all* different possible weights. As before, we have based our approach on the explicit identification of the DEA production possibility set, from which we obtain a complete set of valid weights on which cross-evaluation can be performed. The new approach overcomes existing problems and hence produces more meaningful results. In addition, its rationale can offer an alternative view when imposing restrictions on the self-selection of weights by DMUs in DEA. Finally, we develop a set of extensions and tools to accompany our approach. These allow the identification of so-called maverick DMUs, i.e. DMUs that obtain unrealistic efficiency scores, as well as the identification of under-achieving DMUs, a concept that we introduce here.

Contents

List of Figures

List of Tables

Chapter 1

Measuring efficiency: basic concepts and models

1.1 Introduction

This chapter contains all preliminary material on efficiency measurement/analysis relevant to this thesis. In section 1.2 we consider the problem from an economic viewpoint and introduce some fundamental efficiency concepts. We start by providing the *Pareto-Koopmans* definition of *technical-efficiency* and introduce the concept of the *production possibility set.* We then discuss how *Debreu-Farrel efficiency measures* utilise production possibility sets to measure technical-efficiency and how these differ from Pareto-Koopmans measures. We close this section by discussing the concept of *Returns-to-Scale.* In section 1.3 we introduce *Data Envelopment Analysis* (DEA). We provide a detailed description of how DEA measures efficiency, define the concepts of *relative efficiency* and *radial-efficiency* and present the basic DEA models for *constant* and *variable* returns-to-scale cases. Finally, we discuss *super-efficiency* DEA models and their uses.

1.2 Basic Concepts

1.2.1 Technical Efficiency

The term *production* can be intuitively defined as the utilisation of resources, or more generally inputs, in a process that transforms them into outputs. Usually, this process is taking place under a given technology, assumed to be the same for some production units. Where the production technology is known, e.g. in the form of a production function, it can be used in an attempt to measure the efficiency of production units by comparing their actual input-output bundles with 'optimal' ones implied by the *production technology*, e.g. the ratio of actual to optimal output obtained by a given input. Comparisons such as this measure efficiency in terms of production possibilities and give rise to the term *technical efficiency* (Lovell 1993).

The most widely accepted definition of technical efficiency is provided by Koopmans (1951) and is commonly referred to as *Pareto-Koopmans* efficiency or just *Pareto-efficiency*: a producer is technically efficient if an increase in any output requires a reduction in at least one other output or an increase in at least one input, and if a reduction in any input requires an increase in at least one other input or a reduction in at least one output.

Before illustrating the concept of technical efficiency we present the notation that will be used. Let $J = \{1, 2, ...n\}$ be a set of production units indexed by j. The j'th unit consumes the s-dimensional input vector $\mathbf{x}_j$ to produce the m-dimensional output vector $\mathbf{y}_j$. We also define the $s \times n$ dimensional matrix $\mathbf{X} = \left[\mathbf{x}_1^T, ...\mathbf{x}_n^T\right]$ and the $m \times n$ dimensional matrix $\mathbf{Y} = \left[\mathbf{y}_1^T, ...\mathbf{y}_n^T\right]$, and denote *input-output bundles* in general as $(x, y) \in \mathbb{R}_+^{s+m}$. Finally, we define the production technology in terms of a *production possibility set* T which includes all possible input-output bundles, i.e.:

$$T = \{(x, y) \in \mathbb{R}_+^{s+m} | x \in \mathbb{R}_+^s \text{ can produce } y \in \mathbb{R}_+^m\}$$

We will assume that this set satisfies the following structural properties (see e.g. Fare et al. 1985):

1 'No free-lunch' assumption: If $(0, 0) \in T$ and $(0, y) \in T$ then $y = 0$. In words, this states that production of output is not possible without consumption of some input.

2 T is monotonic: $\forall (x, y) \in T \Rightarrow (x + \alpha, y - \beta) \in T \; \forall \; \alpha \in \mathbb{R}_+^s$, $\beta \in \mathbb{R}_+^m$. This property is also referred to as free disposability of inputs and outputs.

3 T is convex. Let $\mathbf{x}_j$ and $\mathbf{y}_j$ denote the j'th columns of the matrices $\mathbf{X}$ and $\mathbf{Y}$. Then a production set T is convex if: $\forall\ \mathbf{X}, \mathbf{Y} : (\mathbf{x}_j, \mathbf{y}_j) \in T\ \forall j \in J \Rightarrow (\mathbf{X}\lambda, \mathbf{Y}\lambda) \in T :\ \boldsymbol{\lambda} \in \mathbb{R}^n_+,\ e\boldsymbol{\lambda} = 1$ and $e = (1...1)$.

4 T is a closed set.

Perhaps the best known measure of technical efficiency was introduced by Debreu (1951) and Farrell (1957). This measure is defined in terms of the maximal equi-proportionate or *radial* reduction in inputs while outputs remain the same. We refer to this measure as the Debreu-Farrell input oriented measure of technical efficiency and denote it as follows:

$$DF_I(x, y) = \min\{\theta : (\theta x, y) \in T\} \tag{1.1}$$

The Debreu-Farrell measure is closely related to the *input distance function* introduced by Shephard (1953, 1970), defined as follows:

$$D_I(x, y) = \max\{\theta : (x/\theta, y) \in T\} \tag{1.2}$$

We note that $DF_I(x, y) \leq 1$, $D_I(x, y) \geq 1$ and obviously $DF_I(x, y) = 1/D_I(x, y)$. The value of $DF_I(x, y)$ is taken to be the input-oriented efficiency of a given production unit. In an analogous fashion we can define the output oriented versions of the above measures as follows:

$$DF_O(x, y) = \max\{\delta : (x, \delta y) \in T\} \tag{1.3}$$

$$D_O(x, y) = \min\{\delta : (x, y/\delta) :\in T\} \tag{1.4}$$

We note that $DF_O(x, y) \geq 1$, $D_O(x, y) \leq 1$ and in this case we take $1/DF_O(x, y) = D_O(x, y)$ to be the efficiency score so that again all efficiency scores are bounded by unity.

The above measures are in the form of distance functions that measure the distance of the actual input-output bundle of a production unit from a reference point on the graph of the production function. In that sense, the graph of the production function, taken to be the set of all possible efficient input-output bundles, with efficiency scores equal to one, serves as an *efficient frontier.*

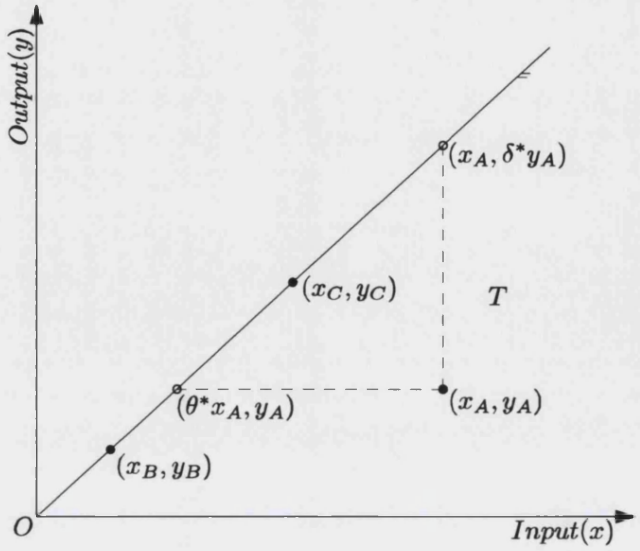


Figure 1.1: Debreu-farrell efficiency

Depending on the orientation, the efficient subset of units consists of those that cannot further decrease (increase) the levels of their inputs (outputs) without affecting the levels of some outputs (inputs). These units achieve an efficiency score equal to one in the corresponding orientation. An illustration is given in figure 1.1. Suppose that the production possibility set is the collection of input-output bundles bounded from the left by the ray from the origin through data points (x_B, y_B) and (x_C, y_C). For units B and C no possible increase in output production is possible without an increase in the levels of input and no possible reduction in input levels is possible without a reduction in output levels. Therefore, units B and C are efficient with efficiency equal to one. Unit A however is not technically efficient since, depending on the orientation, it could have either used less input $(\theta^* x_A)$ to produce the same output, or produced more output $(\delta^* y_A)$ with the same input. The corresponding efficiency scores for unit A are $\theta^*$and $1/\delta^*$ [1].

We need to note here that a unit need not be simultaneously declared efficient by both orientations of the Debreu-Farrell measure, which is a necessary condition[2] for a unit to be Pareto-efficient. More generally, Pareto-efficiency implies Debreu-Farrell efficiency but the converse does not necessarily hold. Fare and Lovell (1978) and Charnes et al. (1978) note that the Debreu-Farrell

[1] In fact, because we assume that the technology only exhibits constant-returns to scale, $\theta^* = 1/\ \delta^*$.

[2] This follows directly from the definition of Pareto-Koopmans efficiency. Note here that this is not a sufficient condition.

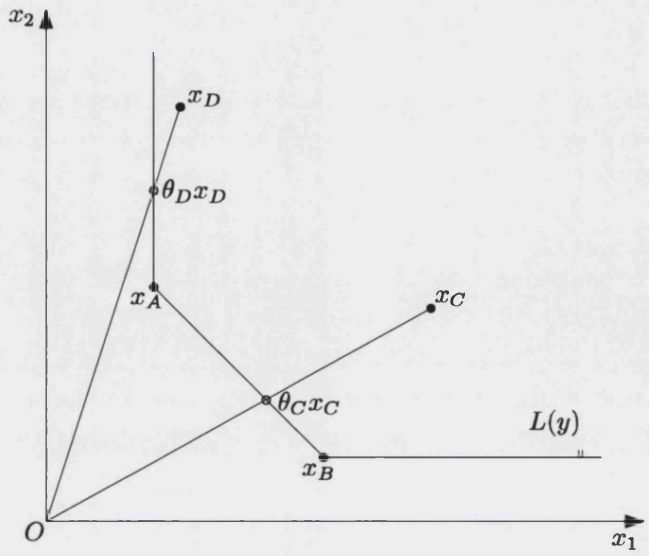


Figure 1.2: Pareto-efficiency vs Debreu-Farrell efficiency

measure of efficiency does not always account for all inefficiency in the Pareto-Koopmans sense. We illustrate such a case in figure 1.2. Consider the plotted *input set* $L(y) \subseteq \mathbb{R}^s_+$, taken to be the collection of input bundles $x \in \mathbb{R}^s_+$ that produce at least output vector $y \in \mathbb{R}^m_+$ (let us suppose here, for simplicity, that all plotted vectors produce exactly the same amount of output)[3]. Both input vectors x_C and x_D can be contracted radially by amount θ_C and θ_D respectively and still be able to produce the same output. Note however, that although no further radial contraction is possible for input vector $\theta_D x_D$, which renders it Debreu-Farrell efficient, this vector is not Pareto-Koopmans efficient since compared to vector x_A it uses more of input x_2 in order to produce the same output. This *slack* in input x_2 results from the inefficient *mix* of inputs that unit D exhibits and is referred to as the *mix*component of inefficiency.

Having observed the relationship between Debreu-Farrell and Pareto-efficiency, we close this section with the identification of the subsets of efficient production bundles associated with these. The *efficient subset* of T denoted $Eff(T)$ includes all Pareto-Koopmans efficient units:

$$Eff(T) = \{(x,y) \in T \mid x' \leq x,\ y' \geq y,\ (x',y') \neq (x,y) \Rightarrow (x',y') \notin T\} \tag{1.5}$$

[3]Note that in an analogous fashion we can define the *output set* $P(x) \subseteq \mathbb{R}^m_+$.

The *weak-efficient* subset of T includes all Debreu-Farrell efficient units (see e.g. Fare et al. 1985; Russell 1988):

$$W.Eff(T) = \{(x,y) \in T \mid x' < x,\ y' > y \Rightarrow (x',y') \notin T\} \tag{1.6}$$

It follows from the above that

$$Eff(T) \subseteq W.Eff(T) \tag{1.7}$$

This 'disagreement' between the two efficiency definitions has always been a weak point of Debreu-Farrell efficiency. Remedies for this have been suggested in the literature (see e.g. Lovell 1993 for a short discussion) and in section 1.3 we will be discussing one such approach.

1.2.2 Returns to Scale

In this section we discuss the important concept of *Returns-to-Scale.* Consider the example in figure 1.3 where the production technology involves one input x and one output y, such that $y = f(x)$. We assume that all pairs (x,y) in this functional relationship are technically efficient, i.e. the output y is maximal for every x. The graph of this function is given by the s-shaped curve and the area to the right of the curve is the production possibility set. The concept of Returns-to-Scale relates to comparing proportional increases in output that result from proportional increases in input, assuming technical efficiency is maintained. Hence, we only concentrate on points that lie on the graph of the production function. Consider the production bundle (x_1, y_1). Its *average productivity* is defined by the ratio y_1/x_1 which gives the slope of the ray from the origin through point (x_1, y_1). If we were to move away from (x_1, y_1) by increasing input, but still remain on the graph, then the slopes of the corresponding rays would increase until point (x_o, y_o), where the ray from the origin to (x_o, y_o) is tangent to the graph, after which they would start decreasing. Hence, average productivity increases along with increases in inputs up to level x_o and decreases after that point. This implies that output is increasing proportionately more than input to the left of x_o and the reverse is observed to the right of x_o.

Formalising this, we say that the production technology exhibits *Increasing Returns-to-Scale* for $x < x_o$, *Constant Returns-to-Scale* at point (x_o, y_o) and *Decreasing Returns-to-Scale* for $x > x_o$.

To generalise our observations, consider a production function $F(x,y) = 0$[4] representing all

[4] This is the standard neoclassical produnction function for multiple inputs and outputs which represents Pareto-

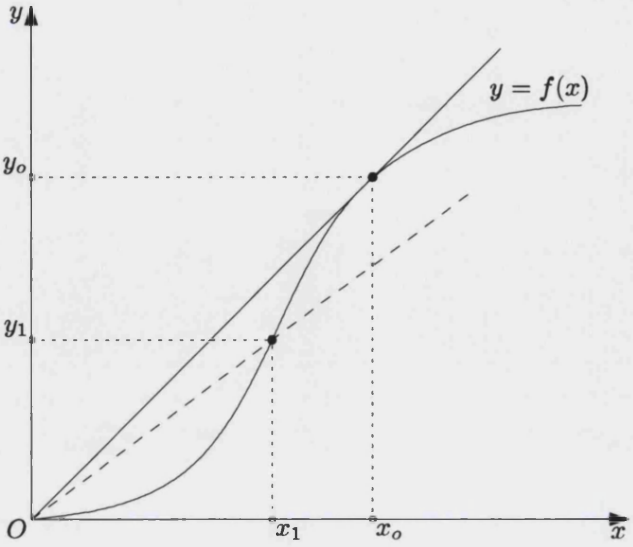


Figure 1.3: Returns-to-Scale and the production function.

Pareto-efficient combinations of inputs and outputs $(x, y) \in \mathbb{R}_+^{s+m}$. We measure the returns-to-scale at a particular production bundle on the graph of the production function, by considering the proportional increases in all output levels relative to a marginal proportional increases in all inputs. More specifically, suppose that inputs are increased proportionally by a factor $\alpha \geq 1$ and consider the proportional expansion in outputs $\beta \geq 1$ required to match this if efficiency is to be maintained, determined by solving $F(\alpha x, \beta y) = 0$. We now define the *elasticity of scale* as a function of inputs and outputs as follows:

$$\varepsilon(x, y) = \frac{\partial \beta}{\partial \alpha} \frac{\alpha}{\beta} \tag{1.8}$$

An alternative formula is obtained by differentiating F with respect to α and evaluating at $\alpha = \beta = 1$ without loss of generality (see e.g. Forsund and Hjalmarsson (2004)):

$$\varepsilon(x, y) = \frac{\partial \beta}{\partial \alpha} = -\frac{\sum_{i=1}^{s} (\frac{\partial F(x,y)}{\partial x_i}) x_i}{\sum_{r=1}^{m} (\frac{\partial F(x,y)}{\partial y_r}) y_r} \tag{1.9}$$

efficient input-output bundles and is assumed to be continuously differentiable. Further, it is assumed that $\frac{\partial F(x,y)}{\partial y_r} > 0 \; \forall r \in (1, ..., m)$ and $\frac{\partial F(x,y)}{\partial x_i} < 0 \; \forall i \in (1, ..., s)$.

In other words, the elasticity of scale evaluated at point (x, y) measures the ratio of a marginal proportional expansion in output levels resulting from a marginal proportional expansion in input levels, when Pareto-efficiency is maintained. If $\varepsilon(x, y) > 1$, then a proportional increase in inputs by α results in a greater proportional increase in output levels by β, i.e. $\alpha < \beta$, and we say that the production technology exhibits Increasing Returns-to-Scale at point (x, y). Conversely, if $\varepsilon(x, y) < 1$, then $\alpha > \beta$ and we say that the production technology exhibits Decreasing Returns-to-Scale at point (x, y). Finally, if $\varepsilon(x, y) = 1$, then $\alpha = \beta$ in which case we have Constant Returns-to-Scale at point (x, y).

1.3 Data Envelopment Analysis

In the previous section we have assumed that the production technology is known. However, in practical problems this is rather uncommon. The task of measuring efficiency when we hold no information on the production technology is considerably more complicated as we now have to estimate it. The first to work on this was Farrell (1957). In his seminal work he pioneered not only the idea of measuring efficiency with the use of distance functions that contract or expand inefficient units to the efficient frontier, as we have seen in the previous section, but also the idea of constructing an estimated frontier by taking combinations of *observed* input-output bundles as *reference* points. Farrel was concerned with the estimation of unit isoquants, for which he also put forward a computational framework. However, Farrell's work on efficiency analysis was "confined to single output cases and his sketch of extensions to multiple outputs did not supply what was required for applications to large data-sets"(Cooper et al. 2004). Charnes, Cooper, and Rhodes (1978) (CCR) extended Farrell's work within a mathematical programming framework by developing the technique which they baptised Data Envelopment Analysis. CCR also introduced the term *Decision Making Units* (DMUs) to refer to a set of peer entities that convert multiple inputs to multiple outputs. In this section we present the basic models of DEA.

1.3.1 Constant Returns-to-Scale models

As a starting point, CCR chose a different side of the problem: the use of input-output data for every DMU in a ratio formula that evaluates efficiency. To achieve this DEA assigns *weights* to

the inputs and outputs and the ratio takes the form: weighted sum of outputs/weighted sum of inputs. Let $v \in \mathbb{R}^s_+$ and $u \in \mathbb{R}^m_+$ be the vectors of weights assigned to the inputs and outputs respectively. The following non-linear programme[5] evaluates the efficiency of $\text{DMU}_O \in J$:

$$\begin{aligned} \max \ & \frac{u\mathbf{y}_o}{v\mathbf{x}_o} \\ s.t. \ & \frac{u\mathbf{y}_j}{v\mathbf{x}_j} \leq 1 \ \forall j \in J \\ & u, v \geq 0 \end{aligned} \tag{1.10}$$

The efficiency of DMU_O given by the optimal value of the objective function in (1.10) is also referred to as *simple-efficiency* because it is the *self-evaluation* of a DMU by choosing weights that place it in the best possible light. In contrast, *cross-efficiency* is the efficiency score obtained by evaluating a DMU with weights that have been selected by *another* DMU in its self-evaluation (we look at cross-efficiency in great detail in chapter 6). In this context, DMU_O is trying to find nonnegative weights that maximise its simple-efficiency, while keeping this and all cross-efficiencies bounded by unity. Efficient DMUs are those that are able to achieve the maximum efficiency score of unity. CCR converted the above non-linear programme into a linear one by applying the *Charnes-Cooper transformation* (Charnes and Cooper 1962). This involves setting the denominator in the objective function equal to one and maximising the numerator[6]. The constraints can easily be rearranged algebraically in a linear form. The CCR Linear Programme (LP) is given below. This is also referred to as the *multiplier form* of the CCR model:

$$\begin{aligned} \max \quad & u\mathbf{y}_o \\ s.t. \quad & v\mathbf{x}_o = 1 \\ & u\mathbf{Y}^T - v\mathbf{X}^T \leq 0 \\ & v, u \geq 0 \end{aligned} \tag{1.11}$$

[5] Note here that to maintain simplicity in presentation we will be denoting u, v instead of u^T, v^T. We will be using this rule in all mathematical notation throughout this text.

[6] Note that by setting $v\mathbf{x}_o = 1$ we assume implicitly that $\mathbf{x}_j \neq \mathbf{0} \ \forall j \in J$.

As before, the optimal value of the objective function gives the efficiency of DMU_O. We note here that a fully rigorous analysis would replace the nonnegativity constraints by the constraints $v, u \geq \varepsilon > 0$, where ε is a *non-Archimedean infinitesimal*[7] We shall use the term *infinitesimal form* to refer to versions of DEA models constructed to accommodate the infinitesimal ε. We discuss this in greater detail later. For now, we turn to the dual to the above LP, known in the literature as the *envelopment form* of a CCR model:

$$\begin{aligned} &\min\ \theta \\ &s.t.\ \theta \mathbf{x}_o - \mathbf{X}\boldsymbol{\lambda} \geq 0 \\ &\mathbf{Y}\boldsymbol{\lambda} \geq \mathbf{y}_o \\ &\boldsymbol{\lambda} \geq 0 \end{aligned} \tag{1.12}$$

We denote the optimal solution to (1.12) by $(\theta^*, \boldsymbol{\lambda}^*)$ and take θ^* to be the input oriented efficiency score for DMU_O. We also refer to this as *radial-efficiency*, as it assumes that all inputs will be contracted by equal proportions. Hence, model (1.12) is an operational form of the Debreu-Farrell input oriented efficiency measure given in (1.1). The point $(\theta^*\mathbf{x}_o, \mathbf{y}_o)$ can be seen as a target point for DMU o, and can also be expressed as a nonnegative linear combination of peer DMUs[8] by using $\boldsymbol{\lambda}^*$. The efficient frontier is the collection of all radially-efficient combinations of DMUs, all of which can serve as target points. It is easy to see that $(\theta = 1,\ \boldsymbol{\lambda}_o = 1)$ is a feasible solution to (1.12) and hence it follows that for all DMUs $\theta^* \leq 1$, implying that no observed input-bundle can be expressed as a radial contraction of a target point. In other words, this model seeks to determine whether there exists a nonnegative combination of DMUs that uses less of every input than DMU o, but produces the same levels of all outputs. If such a combination cannot be found then DMU o is declared radially-efficient. In that sense, the frontier envelopes the observed data and this gives rise to the name DEA. Radially-efficient DMUs are the ones already on the frontier, for which no possible reduction in input levels is possible and $\theta^* = 1$, like DMUs B and C in figure 1.1. This leads to the following definition.

[7] A non-Archimedean infinitesimal is a number that does not satisfy the Archimedean property, i.e. given a real number β, then $\alpha\varepsilon < \beta$ for any natural number α. This means that ε is smaller than any positive real number.

[8] Here we can clearly see the link with Farrell's work on efficiency analysis.

Definition 1 (Weak DEA-Efficiency, Radial-Efficiency) *DMU o is radially-efficient, if and only if at an optimal solution to (1.12) for DMU o,* $\theta^* = 1$.

We have used the term weak DEA-efficiency (or simply weak-efficiency) because radial-efficiency does not necessarily imply Pareto-efficiency[9]. Recall that model (1.12) is an LP implementation of the Debreu-Farrell efficiency measure and hence inherits its weakness related to distinguishing between weak efficiency and true efficiency (in the Pareto-Koopmans sense). In a DEA context, we use the term *DEA-efficient subset* (or simply *efficient-subset*) to refer to all Pareto-efficient units and the term *weak-efficient subset* to refer to the set of all radially-efficient units, i.e. those with efficiency scores equal to one. As the subsequent discussion will show, not all radially-efficient units are DEA-efficient, as for some radially-efficient units slacks in inputs or outputs might be present. These belong to the weak-efficient subset but not the DEA-efficient subset. We use the term *weakly efficient* units to refer to units that only belong to the weak-efficient subset. Before discussing these issues further, we need to stress that the notion of efficiency in DEA has a relative meaning, as it is based on estimated production technologies. This is formally expressed in the following definition (see Charnes and Cooper 1985; Cooper et al. 2004).

Definition 2 (Relative Efficiency, DEA-Efficiency) *A DMU is DEA-efficient on the basis of available evidence if and only if the performances of other DMUs do not show that some of its inputs or outputs can be improved without worsening some other inputs or outputs.*

To properly decide on membership to the DEA-efficient subset for DMU o, we need to modify the objective function in (1.12) to $\theta - \varepsilon(e\mathbf{s}^- + e\mathbf{s}^+)$, where $\mathbf{s}^-$and $\mathbf{s}^+$ contain the input and output slacks respectively. However, solving (1.12) with the modified objective function has its own problems, deriving from the computation of the value of ε. Instead of solving such a model directly, a two-phase approach is used. In the first phase we solve (1.12). We then use the optimal

[9] One might argue against the use of the term Pareto-efficiency in the context of estimated production technologies. However, we choose to use this term for the sake of simplicity and to keep in line with the majority of DEA related literature. Equivalent terms to Pareto-efficiency that appear in the literature and that we will use are: *true efficiciency, full efficiency, strong-efficiency,* or simply *DEA-efficiency.*

θ^* in a second phase where we solve the following additional model:

$$\max es^- + es^+ \tag{1.13}$$

$$s.t.\ \theta^* \mathbf{x}_o - \mathbf{X}\boldsymbol{\lambda} - \mathbf{s}^- = 0 \tag{1.14}$$

$$\mathbf{Y}\boldsymbol{\lambda} - \mathbf{s}^+ = \mathbf{y}_o \tag{1.15}$$

$$\boldsymbol{\lambda},\ \mathbf{s}^-,\ \mathbf{s}^+ \geq 0$$

Let $(\boldsymbol{\lambda}^*, \mathbf{s}^{-*}, \mathbf{s}^{+*})$ be *an* optimal solution[10] to (1.13). We refer to $(\ \theta^*, \boldsymbol{\lambda}^*, \mathbf{s}^{-*}, \mathbf{s}^{+*})$ as a *max-slack* solution for DMU o[11]. Having maximised the slacks we are now able to determine whether a DMU is DEA-efficient. This is done according to the following definition (see e.g. Cooper et al. 2004).

Definition 3 (DEA-Efficiency) *Let* $(\ \theta^*, \boldsymbol{\lambda}^*, \mathbf{s}^{-*}, \mathbf{s}^{+*})$ *be a max-slack solution for DMU* $o \in J$. *Then, DMU o is fully efficient if and only if both (i)* $\theta^* = 1$, *and (ii)* $e\mathbf{s}^{-*} + e\mathbf{s}^{+*} = 0$ [12].

We can also use the max-slack solution to set appropriate *targets* for inefficient and weakly efficient DMUs. Let $(\hat{\mathbf{x}}_o, \hat{\mathbf{y}}_o)$ be the input-oriented target point for DMU o. Then:

$$\begin{aligned} \hat{\mathbf{x}}_o &= \mathbf{X}\boldsymbol{\lambda}^* = \theta^* \mathbf{x}_o - \mathbf{s}^{-*} \\ \hat{\mathbf{y}}_o &= \mathbf{Y}\boldsymbol{\lambda}^* = \mathbf{y}_o + \mathbf{s}^{+*} \end{aligned} \tag{1.16}$$

Peer DMUs used in the construction of the target point are referred to as *comparator DMUs,* and the collection of a DMU's comparators is called the *reference set.* Formally, let REF_o be the reference set for DMU o. Then:

$$REF_o = \{j \in J |\ \boldsymbol{\lambda}_j^* > 0\} \tag{1.17}$$

Where $\boldsymbol{\lambda}_j^*$ is the j'th entry of vector $\boldsymbol{\lambda}^*$.We illustrate these points with the use of two examples. We start with figure 1.4, where seven DMUs are plotted in input-output space. All DMUs use one input (x) to produce one output (y). In single input-output environments, it is easy to see that all DMUs which maximise the ratio: output produced/input used, will be radially-efficient. In our

[10] Note that there might exist multiple optimal solutions for DEA problems.

[11] Let us note here that a max-slack solution can also be obtained by solving the infinitesimal form of (1.12).

[12] This definition is not only restricted to CCR efficiency but extends, as we shall see later, to other DEA models. Note that condition ii implies that $\mathbf{s}^{-*} = \mathbf{s}^{+*} = 0$.

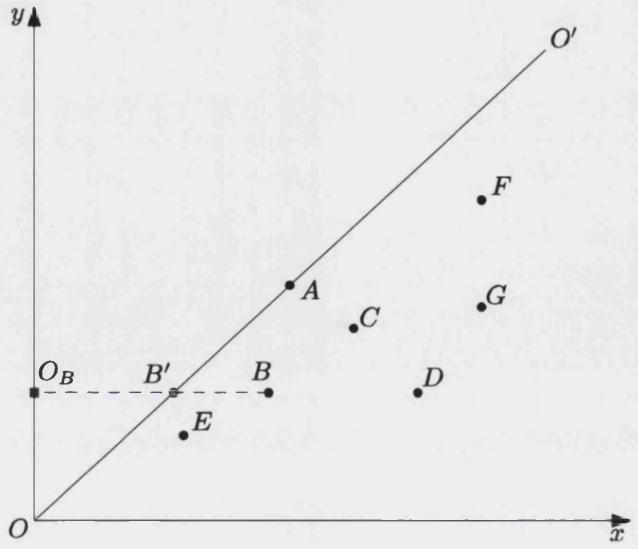


Figure 1.4: CRS example, input-output space

example this ratio is maximised at DMU A only and therefore A is the only efficient DMU. The efficient frontier is given by the ray OO' which includes all input-output bundles that obtain an equal value to DMU A in the above ratio. Of course, all these can be expressed as nonnegative multiples of A. The remaining DMUs are inefficient. We obtain their efficiency scores by projecting their input-output bundles horizontally to the frontier, so that their input levels decrease but their output-levels remain the same. For instance, in the context of (1.12) DMU B's efficiency score would be $\theta_B^* : (\theta_B^* x_B, y_B) = \boldsymbol{\lambda}^*(x_A, y_A) = (x_{B'}, y_{B'})$. Note that $\theta_B^* x_B = x_{B'} \Leftrightarrow \theta_B^* = \frac{x_{B'}}{x_B}$, or else $\theta_B^* = \frac{|O_B B'|}{|O_B B|}$ and hence we can also express efficiency scores with the use of the ratio: *optimal input levels/observed input levels.*

The example discussed above gives a graphical illustration of DEA in the full input-output space, but does not allow for illustrating many important aspects of DEA. For this purpose we will use an additional example in input-space only. In figure 1.5[13], we assume that all plotted DMUs use different levels of two inputs to produce exactly one unit of one output. We start with applying the first phase of the DEA optimisation procedure, i.e. solving (1.12) for all DMUs. At this stage we are concerned with determining the maximum radial contraction in the input levels of a DMU such that the resulting input bundle can be expressed as a combination of other observed bundles. In

[13] For simplicity, x_j refers to (x_{1j}, x_{2j})

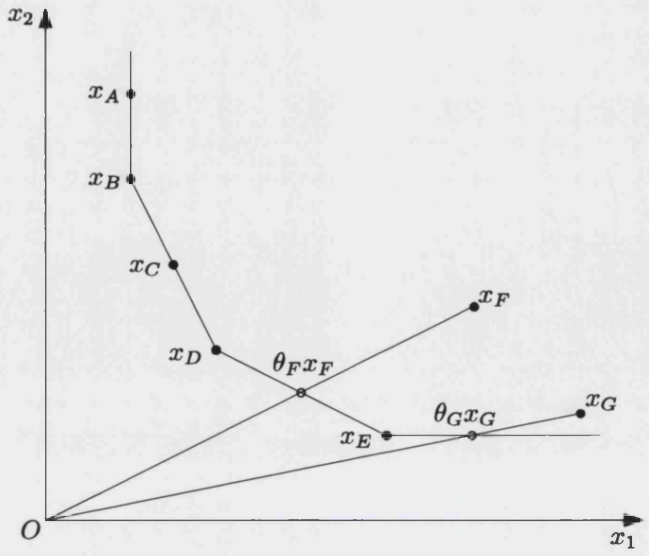


Figure 1.5: CRS example, input space

other words, phase-1 identifies the radial-efficiency scores for all DMUs. DMUs $A-E$ are radially-efficient as there can be no radial contraction of their input levels. DMUs F and G however, can be contracted radially to points $\theta_F x_F$ and $\theta_G x_G$ and hence are inefficient with efficiency scores θ_F and θ_G respectively. As before, we can alternatively express the efficiency scores with use of ratios, e.g. $\frac{|O\theta_F x_F|}{|Ox_F|}$.

After obtaining the radial-efficiencies we need to identify the correct slacks and targets for DMUs where appropriate. This cannot be done by using the results of phase-1. Note the existence of alternative optimal solutions for DMUs A and C. Without affecting its radial-efficiency score, DMU C can either choose itself as comparator or it can choose to be expressed as a combination of B and D. Similarly, DMU A can either chose itself or it can be compared with DMU B. However, for DMU A, such a choice has important implications. As we can observe, choosing DMU B as comparator results in a slack in input x_2 (equal to $x_{2A}-x_{2B}$), which is not the case if it chooses to be compared with itself. Note also, the slack in input x_1 for DMU G ($\theta_G x_{1G}-x_{1E}$). By maximising the slacks, phase-2 guarantees that we arrive at appropriate results so that we can distinguish with certainty between fully efficient DMUs ($B-E$) and weakly efficient DMUs (A). In addition we can set appropriate targets for weakly efficient DMUs and inefficient DMUs.

We close this section with a brief discussion on the output oriented versions of the CCR models (1.11) and (1.12). Note that we concentrate on the radial-efficiency versions of the models, i.e. the models solved in phase-1. The dual pair of models are as follows:

$$\begin{aligned} \max \quad & \varphi \\ s.t. \quad & \mathbf{X}\boldsymbol{\lambda} \leq \mathbf{x}_o \\ & \varphi \mathbf{y}_o - \mathbf{Y}\boldsymbol{\lambda} \leq 0 \\ & \boldsymbol{\lambda} \geq 0 \end{aligned} \tag{1.18}$$

$$\begin{aligned} \min \quad & v\mathbf{x}_o \\ s.t. \quad & u\mathbf{y}_o = 1 \\ & v\mathbf{X}^T - u\mathbf{Y}^T \geq 0 \\ & v, u \geq 0 \end{aligned} \tag{1.19}$$

We denote by $(\varphi^*, \boldsymbol{\lambda}^*)$ the optimal solution to (1.18). It is easy to see that $(\varphi = 1, \lambda_o = 1)$ is a feasible solution, and so $\varphi^* \geq 1$. The output-oriented radial-efficiency score for DMU o is defined as $\frac{1}{\varphi^*}$. All efficient units have $\varphi^* = 1$. Finally, let (v^*, u^*, β^*) be an optimal solution to (1.19). Then, from duality, $v^*\mathbf{x}_o = \varphi^*$ so that $\frac{1}{v^*\mathbf{x}_o}$ provides the output-oriented radial-efficiency score for DMU o.

1.3.2 Variable Returns-to-Scale models

So far we have discussed models that are used in Constant Returns-to-Scale (CRS) environments. Banker, Charnes, and Cooper (1984)(BCC) introduced the following DEA model that can account

for Variable Returns-to-Scale (VRS):

$$\begin{aligned} &\min \theta - \varepsilon(e\mathbf{s}^- + e\mathbf{s}^+) \\ &s.t.\ \theta\mathbf{x}_o - \mathbf{X}\boldsymbol{\lambda} - \mathbf{s}^+ = 0 \\ &\quad \mathbf{Y}\boldsymbol{\lambda} - \mathbf{s}^- = \mathbf{y}_o \\ &\quad e\boldsymbol{\lambda} = 1 \\ &\quad \boldsymbol{\lambda}, \mathbf{s}^+, \mathbf{s}^- \geq 0 \end{aligned} \tag{1.20}$$

Let ($\theta^*, \boldsymbol{\lambda}^*, \mathbf{s}^{-*}, \mathbf{s}^{+*}$) denote an optimal solution to the above LP. The input-oriented radial-efficiency score for DMU o is θ^*. It is easily noticeable that this LP is the same as the envelopment form of the CCR model with the exception of an additional constraint. Constraint $e\boldsymbol{\lambda} = 1$, also referred to as the *convexity constraint*, implies that reference points (and the efficient frontier) are constructed as convex combinations of observed data points. As in the CRS case, instead of solving (1.20) directly we can employ a two-phase procedure. Given a max-slack solution for (1.20), the same definitions for radial-efficiency and DEA-efficiency, given before, apply. In addition target points can be identified as in (1.16).

The multiplier form of this Variable Returns to Scale (VRS) model is given below. Note that the difference with the CRS model in multiplier form (1.11) is the additional variable β which is unrestricted in sign.

$$\begin{aligned} \max \quad & u\mathbf{y}_o - \beta \\ s.t. \quad & v\mathbf{x}_o = 1 \\ & u\mathbf{Y}^T - v\mathbf{X}^T - \beta e \leq 0 \\ & u, v \geq \varepsilon \end{aligned} \tag{1.21}$$

Let (v^*, u^*, β^*) be an optimal solution to (1.21). The optimal value of the objective function ($u^*\mathbf{y}_o - \beta^*$) is equal to the optimal value of the objective function in (1.20) and hence also gives the radial-efficiency score for DMU o.

Below we present the output oriented versions of BCC models (1.20) and (1.21). For the sake

of simplicity we concentrate on models that only evaluate radial-efficiency scores, i.e. disregard any possible slacks. The dual pair of models is given below.

$$\begin{aligned} \max \quad & \varphi \\ s.t. \quad & \mathbf{X}\boldsymbol{\lambda} \leq \mathbf{x}_o \\ & \varphi \mathbf{y}_o - \mathbf{Y}\boldsymbol{\lambda} \leq 0 \\ & e\boldsymbol{\lambda} = 1 \\ & \boldsymbol{\lambda} \geq 0 \end{aligned} \tag{1.22}$$

$$\begin{aligned} \min \quad & v\mathbf{x}_o - \beta \\ s.t. \quad & u\mathbf{y}_o = 1 \\ & v\mathbf{X}^T - u\mathbf{Y}^T - \beta e \geq 0 \\ & v, u \geq 0 \end{aligned} \tag{1.23}$$

Let $(\varphi^*, \boldsymbol{\lambda}^*)$ be an optimal solution to (1.22). Clearly then, $\varphi^* \geq 1$. As for the CRS case, the output-oriented radial-efficiency score for DMU o is defined as $\frac{1}{\varphi^*}$, and all efficient units have $\varphi^* = 1$. Let (v^*, u^*, β^*) be an optimal solution to (1.23). Then $v^*\mathbf{x}_o - \beta^* = \varphi^*$ so that the output-oriented radial-efficiency score for DMU o is also given by $\frac{1}{v^*\mathbf{x}_o - \beta^*}$.

In figure 1.6 we provide an illustration of the DEA models where we contrast VRS and CRS results. The horizontal projections refer to input-oriented models (superscript I), the vertical projections to output oriented models (superscript O) and we have used the subscripts C and V, to refer to CRS and VRS projections respectively. For example, F_V^I is the input-oriented VRS reference point for DMU F. The VRS efficient frontier is given by the piecewise linear segment $ABCD$. Although radial DEA projections on the line segment AA' and the ray from E towards $(+\infty, y_E)$ (through H) are indeed possible, these sections are not part of the VRS efficient frontier, as such projections always have non-zero slacks. We will refer to these sections as the weakly efficient parts of the DEA frontier. The CRS efficient-frontier is the ray OO'. The two frontiers intersect at the line segment BC, hence DMUs B and C are both CRS and VRS efficient, whereas

DMUs A, D and E are VRS efficient only. In addition all VRS efficiency scores are clearly greater than, or equal to the corresponding CRS efficiency scores. DMU F for instance has $\frac{|O_F^I F_V^I|}{|O_F^I F|} > \frac{|O_F^I F_C^I|}{|O_F^I F|}$, and equal output oriented CRS and VRS efficiency scores ($\frac{|O_F^O F|}{|O_F^O F^o|}$). Another important difference of the VRS model is that input and output oriented efficiency scores are not necessarily equal as was the case for CRS models. Moreover, it is possible that a DMU is radially-efficient with respect to a certain VRS orientation and inefficient with respect to the 'opposite' orientation, as for example DMU H which is radially-efficient with respect to an output orientation but inefficient with respect to an input orientation. Taking this further, we can say that it is not necessarily the case that both input and output-oriented radial targets are DEA-efficient.

It follows from Definition (2) that radial-efficiency equal to one with respect to both orientations of a CRS (or a VRS) model is a necessary condition for DEA-efficiency. On the other hand, this condition is not sufficient because equi-proportionate contractions or expansions in input or output levels cannot account for slacks in inputs and/or outputs. In general, if $Eff(J)$ denotes the subset of DMUs that are Pareto-efficient and $W.Eff_I(J)$ (respectively $W.Eff_O(J)$) denotes the subset of DMUs that are declared radially efficient with respect to an input (output) orientation, then the following relation holds (see also Krivonozhko et al. 2005):

$$Eff(J) \subseteq W.Eff_I(J) \cap W.Eff_O(J)$$

1.3.3 Super-Efficiency models

We close this chapter with a discussion on another important family of DEA models referred to as *super-efficiency* models, that have been used to rank DMUs and identify outliers. These models first appeared in Banker and Gifford (1988), Banker et al. (1989) and also in Andersen and Petersen (1993) where the term super-efficiency was introduced. Consider the input-oriented

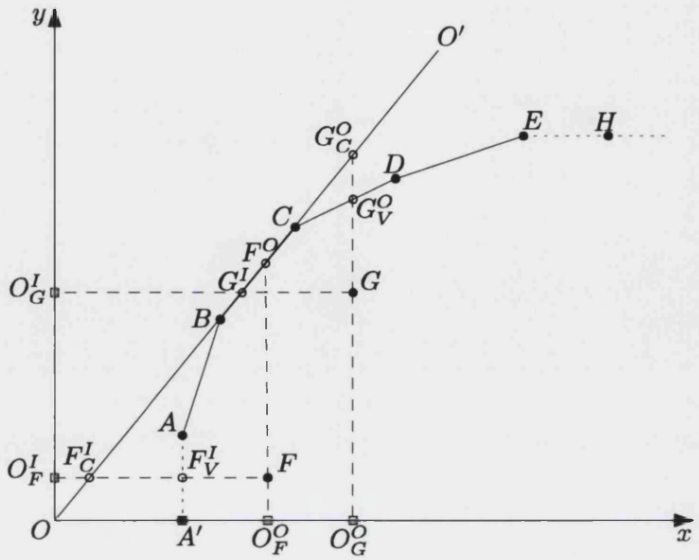


Figure 1.6: CRS vs VRS DEA efficiencies

VRS super-efficiency model given below:

$$\begin{aligned} &\min \theta && (1.24)\\ &s.t.\ \theta \mathbf{x}_o - \mathbf{X}_{-o}\boldsymbol{\lambda} \geq 0 \\ &\quad \mathbf{Y}_{-o}\boldsymbol{\lambda} \geq \mathbf{y}_o \\ &\quad e\boldsymbol{\lambda} = 1 \\ &\quad \boldsymbol{\lambda} \in \mathbb{R}_+^{n-1} \end{aligned}$$

where the matrices $\mathbf{X}_{-o}$ and $\mathbf{Y}_{-o}$ contain all columns $\mathbf{x}_j, \mathbf{y}_j$ except for $\mathbf{x}_o, \mathbf{y}_o$. In other words, λ_o is effectively set to zero and DMU o is excluded from its own reference set. The implications of this are discussed below.

Let the optimal solution to (1.24) be denoted $(\theta', \boldsymbol{\lambda}')$. Now suppose that DMU o is inefficient under the standard BCC model and let the optimal solution to this model instance be denoted $(\theta^*, \boldsymbol{\lambda}^*)$. In such a case DMU o cannot choose itself as comparator in the BCC model, i.e. $\lambda_o^* = 0$. Hence, any optimal solution to (1.24) is an optimal solution to the BCC model and vice-versa (note that the optimal solutions are not necessarily identical as there might exist multiple optima

in both models).

Now suppose that DMU o is a BCC-efficient DMU. We can now further distinguish between the following two cases: a) the only solution to (1.20) (disregarding slacks in the objective) involves $\lambda_o = 1$ and $\lambda_j = 0\ \forall j \neq o$, and b) in addition to the solution in a) there exists at least one other optimal solution to (1.20) in which $\lambda_o = 0$. In the second case, any solution to (1.20) in which $\lambda_o = 0$ is also an optimal solution to model (1.24) and vice-versa. However, in the first case, the optimal solution to (1.20) is no longer feasible for model (1.24). This implies that $\theta' > \theta^* = 1 \Rightarrow\ \theta' > 1$.

To get an alternative interpretation to this consider the dual LP to (1.24), given below:

$$\begin{aligned} \max \quad & u\mathbf{y}_o - \beta \\ s.t. \quad & v\mathbf{x}_o = 1 \\ & u\mathbf{Y}_{-o}^T - v\mathbf{X}_{-o}^T - \beta e \leq 0 \\ & u, v \geq 0 \end{aligned} \tag{1.25}$$

It is evident by comparing this formulation to the multiplier form of the BCC model that the constraint $-v\mathbf{x}_o + u\mathbf{y}_o - \beta \leq 0$ has now been lifted. In addition, if $-v\mathbf{x}_o + u\mathbf{y}_o - \beta > 0$ then, since $v\mathbf{x}_o = 1$, we obtain that $u\mathbf{y}_o - \beta > 1$, i.e. the efficiency score for DMU o is greater than one.

Let us note here that there exist problems regarding the computation of super-efficiency models, stemming from the fact that the feasibility of these models is not guaranteed. An obvious example of an infeasible instance for (1.24) is the instance corresponding to a DMU k such that $y_{1k} > y_{1j}\ \forall j \neq k$, i.e. a DMU k that produces the greatest level of output y_1. Clearly, in such a case y_{1k} cannot be expressed as a convex combination of any y_{1j} values leading to infeasibility in model (1.24). Computational issues arising from the infeasibility of super-efficiency models are discussed in Seiford and Zhu (1999) and Dula and Hickman (1997). Despite these computational challenges, many solutions have appeared in the related literature (see e.g. Lovell and Rouse 2003; Banker and Chang 2006) and the use of super-efficiency models as modelling tools has spread widely.

More specifically, because super-efficiency models do not restrict efficiency scores to be bounded by unity, it has been suggested that they can be used for the following two purposes:

a) Ranking DMUs (Andersen and Petersen 1993): Standard DEA efficiency scores do not allow

for a complete ranking of DMUs. The reason is that all efficient DMUs have equal efficiency scores. By considering the super-efficiency scores this problem is alleviated and a ranking of all DMUs can be constructed.

b) Detecting outliers (Banker and Gifford 1988; Banker et al. 1989): Given that DEA is attempting to fit a frontier/envelope to a set of data points, it is to be expected that it would be sensitive to the presence of DMUs with unrealistically high/low output/input levels (on the other hand it is not sensitive at all to the presence of DMUs with unrealistically high/low input/output levels). In that sense outlier DMUs could be identified as those that are considerably further away from the majority of their peers, in directions indicated by decreases/increases in input/output levels. This 'distance' can be measured by considering the difference between super-efficiency scores and standard DEA efficiency scores (given that all outlier DMUs would have to be radially-efficient, we can identify outliers by just considering their super-efficiency scores). In view of this, the procedure suggested by Banker and Gifford (1988) is based on choosing a threshold value (say 1.1) and identifying as outliers all DMUs with super-efficiency scores that exceed this threshold.

Recently, Banker and Chang (2006) conducted simulation experiments to assess the performance of super-efficiency models for the aforementioned purposes. They report that the outlier identification process discussed above performs satisfactorily but that the use of super-efficiency models in ranking DMUs is not appropriate.

Finally, let us note that the super-efficiency models have been widely used in the sensitivity and stability analysis of DEA efficiency results. In chapter 4 we shall report on this use in more detail.

Chapter 2

On DEA production possibility sets

2.1 Introduction

In the previous chapter we have seen that DEA models construct estimated production technologies by identifying efficient combinations of observed input-output bundles. Our discussion was based at the DMU-level; we presented the theory behind the efficiency classifications of DMUs but we did not examine in detail the specific characteristics of the estimated production technologies. In this chapter we take a macroscopic view of these production technologies. More specifically, we provide a detailed examination of the theoretical framework regarding the various production possibility sets associated with different DEA models.

The outline of this chapter is as follows. In section 2.2 we discuss a selection of topics from linear algebra and polyhedral theory. We have selected the contents of this section so that they form a concise background behind the theory of polyhedral production possibility sets which are the focus of the rest of this chapter as well as subsequent chapters. Section 2.3 introduces two alternative representations of DEA production possibility sets. We provide the theory relating to *primal* and *dual* representations, as well as some illustrative examples. Finally, because a significant amount of the ideas in subsequent chapters is based on the explicit identification of

DEA production possibility sets, in section 2.4 we review what has already been reported in the literature on this.

2.2 Basic concepts

This section presents some basic concepts from *linear algebra, convex analysis* and *polyhedral theory*. Our purpose here is not to provide a comprehensive introduction to these fields but to collect and present the underlying theory for the discussion contained in subsequent sections. For this reason we shall keep this section brief.

2.2.1 Linear algebra

Consider a set of points $S \subseteq \mathbb{R}^n$ and a subset $X \subseteq S$ that contains a finite collection of t points from S, i.e. $X = \{x_1, ..., x_t\}$. Using this we provide the following fundamental definition:

Definition 4 *Consider the point* $x^* \in \mathbb{R}^n$ $x^* = [x_1, ..., x_t]\boldsymbol{\lambda}$. *Then* x^* *is:*

a) a linear combination of points in X *if:* $\boldsymbol{\lambda} \in \mathbb{R}^t$.

b) a conic (or non-negative linear) combination of points in X *if:* $\boldsymbol{\lambda} \in \mathbb{R}^t_+$.

c) an affine combination of points in X *if:* $\boldsymbol{\lambda} \in \mathbb{R}^t$ *and* $e\boldsymbol{\lambda} = 1$.

d) a convex combination of points in X *if:* $\boldsymbol{\lambda} \in \mathbb{R}^t_+$ *and* $e\boldsymbol{\lambda} = 1$.

Now consider the set S again. The *convex hull of* S is defined as the set of all convex combinations of points in S. Alternatively we can define this as the smallest *convex set* that contains S, or even as the intersection of all convex sets that contain S. We denote the convex hull of S by $Conv(S)$. In parallel we can define the *linear hull*, denoted $Lin(S)$, the *affine hull*, denoted $Aff(S)$, and the *conical hull*, denoted $Cone(S)$. In view of the definition above, it is easy to observe that the following relations hold:

$$\begin{aligned} Conv(S) \subseteq Aff(S) \subseteq Lin(S) \\ Conv(S) \subseteq Cone(S) \subseteq Lin(S) \end{aligned} \tag{2.1}$$

In relation to the above we can define the concepts of *linear/affine independence* as follows:

Definition 5 *A set of t vectors* $x_1, ..., x_t$ *are*

a) linearly independent if the unique solution to $[x_1, ..., x_t]\boldsymbol{\lambda} = 0$ *is* $\boldsymbol{\lambda} = \mathbf{0}$.

b) affinely independent if the unique solution to $[x_1, ..., x_t]\boldsymbol{\lambda} = 0$, $e\boldsymbol{\lambda} = 0$ *is* $\boldsymbol{\lambda} = \mathbf{0}$.

An alternative view of the above definition would be to declare the set of vectors $X = \{x_1, ..., x_t\}$ as linearly (respectively affinely) independent if $\nexists x_o \in X : x_o \in Lin(X \setminus x_o)$ (respectively $Aff(X \setminus x_o)$). It is easy to see that linear independence implies affine independence but the converse is not necessarily true.

The maximum number of linearly independent vectors in $\mathbb{R}^n$ is n, whereas the maximum number of affinely independent vectors in $\mathbb{R}^n$ is $n+1$ (e.g. n linearly independent points and the zero vector).

We now turn to the issue of *dimension.* To address this it is necessary to define *linear subspaces* and *affine sets.*

Definition 6 *Let* V *be a subset of* $\mathbb{R}^n$. *Then* V *is a subspace if and only if:*

a) $x_1, x_2 \in V \Rightarrow x_1 + x_2 \in V$.

b) $x \in V \Rightarrow \lambda x \in V \ \forall \lambda \in \mathbb{R}$.

Note that property b in (6) implies that V is not empty as it always contains the zero vector. If $V \neq \mathbb{R}^n$ we call V a proper subspace of $\mathbb{R}^n$. Now suppose that $V \neq \{\mathbf{0}\}$ and consider a finite collection of linearly independent vectors $X = \{x_1, ..., x_t\} \subseteq V$ such that $Lin(X) = V$. Then X is called a basis of V. Bases are not unique, however they all contain the same number of vectors. This number is called the *dimension* of the subspace V and denoted $dim(V)$. Clearly then $\mathbb{R}^n$ has dimension n, and $dim(V) < n$ for every proper subspace V of $\mathbb{R}^n$. If the maximum number of linearly independent points in subspace V is k then the maximum number of affinely independent points in V is $k + 1$. In general, if t_l is the maximum number of linearly independent points in V and t_a is the maximum number of affinely independent points in V, then:

$$dim(V) = t_l = t_a - 1. \tag{2.2}$$

We now discuss another way of determining the dimension of subspaces related to the alternative definition of a subspace below [1]:

[1] Some of the theoretical properties that we present in this section as definitions, are in fact fundamental provable results in linear algebra and convex analysis. Presenting them as such however is beyond the scope of this short

Definition 7 *Let V be a subset of $\mathbb{R}^n$. Then V is a subspace if an only if $\exists$ an $m \times n$ matrix $\mathbf{A}$ such that $V = \{x \in \mathbb{R}^n | \mathbf{A}x = \mathbf{0}\}$.*

In other words the subspace V is the solution set of a finite number of linear equalities of the form $\mathbf{a}x = 0$, given by the rows of $\mathbf{A}$. The maximum number of linearly independent rows of $\mathbf{A}$ is called the *rank* of $\mathbf{A}$ and denoted $rank(\mathbf{A})$. Clearly, $rank(\mathbf{A}) \leq n$. Given $V = \{x \in \mathbb{R}^n | \mathbf{A}x = \mathbf{0}\}$, and a matrix $\bar{\mathbf{A}}$ containing $\bar{m} = rank(\mathbf{A})$ linearly independent rows of $\mathbf{A}$, then $V = \{x \in \mathbb{R}^n | \bar{\mathbf{A}}x = \mathbf{0}\}$ and the latter is a *minimal* description *of* V. The dimension of V is closely associated with the number of equalities required for its minimal description in other words the rank of matrix $\mathbf{A}$. In particular, the subspace V loses one dimension for every one of the linearly independent equalities required for its minimal description, i.e.:

$$dim(V) = n - rank(\mathbf{A}). \tag{2.3}$$

Now consider affine sets in $\mathbb{R}^n$ for which we use the following definition:

Definition 8 *A set $S \subseteq \mathbb{R}^n$ is affine if and only if $\lambda x_1 + (1 - \lambda)x_2 \in S$ for any $x_1, x_2 \in S$. In addition, if $S \subseteq \mathbb{R}^n$ is an affine set then it is the translation of a unique subspace $V \subseteq \mathbb{R}^n$, i.e. $S = V + \mathbf{a} = \{x + \mathbf{a} : x \in V\}$, $\mathbf{a} \in \mathbb{R}^n$.*

Clearly, the affine hull of a set of points is an affine set. Conversely, every affine set $S \subseteq \mathbb{R}^n$ is the affine hull of a set of points in $\mathbb{R}^n$. In addition, for every affine set $S \subseteq \mathbb{R}^n$ there exists an $m \times n$ matrix $\mathbf{A}$ and a vector $\mathbf{b} \in \mathbb{R}^n$ such that $S = \{x \in \mathbb{R}^n | \mathbf{A}x = \mathbf{b}\}$. In other words, the affine set S is the set of solutions to a finite set of equalities of the form $\mathbf{a}x = b$, defined by the rows of $\mathbf{A}$ and $\mathbf{b}$. Any subspace is an affine set and in fact subspaces are those affine sets that contain the origin.

If $S \subseteq \mathbb{R}^n$ is an affine set and V its corresponding subspace then the dimension of S is:

$$dim(S) = dim(V) = n - rank(\mathbf{A}). \tag{2.4}$$

introduction. For a detailed exposition to these topics the reader can refer to Rockafellar (1970), Nemhauser and Wolsey (1988) and Schrijver (2000).

Recall that the dimension of V is equal to one minus the maximum number of affinely independent points in V. Now consider a maximum collection of t affinely independent points $x_1, ..., x_t \in V$. Since S is a translation of V by some $\mathbf{a} \in \mathbb{R}^n$, then the points $x_i + \mathbf{a}$, $i = 1, ...t$, are contained in S and can be seen to be affinely independent. If follows that if t_s is the maximum number of affinely independent points in S, then:

$$dim(S) = t_s - 1. \tag{2.5}$$

Although both linear and affine independence can be used to determine the dimension of subspaces, the same is not true for affine sets and this is why we determine the dimension of these by considering the dimension of their corresponding subspace. For example, consider the subspace $V_1 \subseteq \mathbb{R}^2$ defined by the line $x - y = 0$. The maximum number of linearly independent points in V_1 is clearly one. Now consider the affine set $S_1 = V_1 + (1, 0) = \{(x, y) \in \mathbb{R}^2 | \ x - y = 1\}$. The maximum number of linearly independent points in S_1 is two (e.g. the points $(1, 0)$ and $(0, -1)$). The dimension of S_1 however is $dim(S_1) = dim(V_1) = 1$. By using affine independence we obtain that V_1 contains at most two affinely independent points (e.g. point $(1, 1)$ and the origin) and hence $dim(V) = 2 - 1 = dim(S)$. Note that in general if an affine set S does not contain the origin (i.e. if it is not a subspace), then the maximum number of linearly independent points in S equals the maximum number of affinely independent points in S (e.g. the set S_1 and the points $(1, 0)$ and $(0, -1)$ discussed before).

The maximal affine sets not equal to the whole space $\mathbb{R}^n$ are of particular importance. These are called *hyperplanes.* To be precise, a hyperplane in $\mathbb{R}^n$ is an affine set with dimension $n - 1$. From definition (2.3) we know that exactly one linear equality is required to describe a hyperplane in $\mathbb{R}^n$, i.e.:

Definition 9 *A hyperplane $H \subseteq \mathbb{R}^n$ is defined as: $H = \{x \in \mathbb{R}^n | \ \mathbf{a}x = b, \ \mathbf{a} \in \mathbb{R}^n, \ b \in \mathbb{R}\}$. This representation is unique up to scalar multiplication.*

Since every affine set $S \subseteq \mathbb{R}^n$ is the solution set of a finite number of linear equalities, it follows that $S \subseteq \mathbb{R}^n$ is the intersection of a finite number of hyperplanes.

2.2.2 Polyhedral theory

We now discuss systems of linear inequalities and *polyhedra.* Earlier we defined a hyperplane H as the solution set of one linear equality. Now we define a *halfspace* $H^{\leq}$ as the solution set of a linear inequality[2], i.e. $H^{\leq} = \{x \in \mathbb{R}^n |\ \mathbf{a}x \leq b,\ \mathbf{a} \in \mathbb{R}^n,\ b \in \mathbb{R}\}$. Clearly $H \subset H^{\leq}$. Taking this further, in parallel to affine sets defined as the solution sets for a system of linear equalities we can define a polyhedron as the solution set for a number of linear inequalities, i.e.:

Definition 10 *A polyhedron (or a polyhedral set) $P \subseteq \mathbb{R}^n$ is the set of solutions to a system of linear inequalities. More specifically, If P is a polyhedron, then there exist an $m \times n$ matrix $\mathbf{A}$ and a vector $\mathbf{b} \in \mathbb{R}^m$ such that $P = \{x \in \mathbb{R}^n | \mathbf{A}x \leq \mathbf{b}\}$.*

A special case of polyhedra are *polyhedral cones*[3], for which the vector $\mathbf{b}$ is equal to the zero vector $\mathbf{0}$, i.e. if $C \subseteq \mathbb{R}^n$ is a polyhedral cone, then $C = \{x \in \mathbb{R}^n | \mathbf{A}x \leq \mathbf{0}\}$ Another special case is *bounded polyhedra* which are also called *polytopes.* More specifically, if $\Pi \subseteq \mathbb{R}^n$ is a polytope then $\Pi \subseteq \{x \in \mathbb{R}^n | -\boldsymbol{\delta} \leq x \leq \boldsymbol{\delta},\ \boldsymbol{\delta} \in \mathbb{R}^n_+\}$.

Note that two inequalities $\mathbf{a}_o x \leq b_o$ and $\mathbf{a}_o x \geq b_o$ together are equivalent to the equality $\mathbf{a}_o x = b_o$, so that *implied equalities* might be present in the definition of polyhedron P. In general, let $\mathbf{A}^{=}x = \mathbf{b}^{=}$ denote a set of m_1 linear equalities, $\mathbf{A}^{\leq}x \leq \mathbf{b}^{\leq}$ a set of m_2 linear inequalities and let $m = m_1 + m_2$. Then any polyhedron $P \subseteq \mathbb{R}^n$ can be described as $P = \{x \in \mathbb{R}^n | \mathbf{A}^{=}x = \mathbf{b}^{=}, \mathbf{A}^{\leq} x \leq \mathbf{b}^{\leq}\}$. The dimension of a polyhedron P is closely related to the number of linear equalities required for its description. In parallel to our earlier discussion we define the dimension of $P \subseteq \mathbb{R}^n$ as follows:

$$dim(P) = n - rank(\mathbf{A}^{=}, \mathbf{b}^{=}) \tag{2.6}$$

Consider a point $x_o \in P$ such that $\mathbf{A}x < \mathbf{b}$. Such a point is called an *interior point.* Clearly then a polyhedron $P \subseteq \mathbb{R}^n$ is of full dimension ($dim(P) = n$) if and only if it has an interior point. In addition, we can determine the dimension of a polyhedron by considering the maximum number of affinely independent points that it contains. Let that number be t. Then:

$$dim(P) = t - 1 \tag{2.7}$$

[2] Without loss of generality we assume that all inequalities are of 'less than or equal to' form.
[3] $C \subseteq \mathbb{R}^n$ is a cone if: $x \in C \Rightarrow \lambda x \in C,\ \lambda \in \mathbb{R}_+$.

In subsequent sections and chapters we will only encounter fully-dimensional polyhedra so we shall restrict the remaining parts of this section to this special case. We now come to the issue of the minimal representation of P. Consider an inequality $\mathbf{a}_o x \leq b_o$ satisfied by every $x \in P$. Such an inequality is called a *valid inequality* and we say that $(\mathbf{a}_o, b_o)$ defines the inequality. Clearly the valid inequalities for P are infinite, however, not all valid inequalities are required for the description of P. Furthermore, even the finite system $\mathbf{A}x \leq \mathbf{b}$ might contain inequalities that are not required for representing P. To obtain the minimal representation of P we need to consider the *faces* and *facets* of P, defined as follows:

Definition 11 *Let $(\mathbf{a}_o, b_o)$ define a valid inequality for a polyhedron $P \subseteq \mathbb{R}^n$, and $f = \{x \in P | \mathbf{a}_o x = b_o\}$. Then, f is called a face of P and we say that $(\mathbf{a}_o, b_o)$ represents f. If $f \neq \varnothing$ then we also say that $(\mathbf{a}_o, b_o)$ supports P and call the hyperplane $H = \{x \in \mathbb{R}^n | \mathbf{a}_o x = b_o\}$ a supporting hyperplane of P (and f). Moreover, we call f a proper face if $f \neq \varnothing$ and $f \neq P$. .*

Note that the number of distinct faces of P is finite. In addition we define facets as follows:

Definition 12 *Let f be a face of P. If $dim(f) = dim(P) - 1$ then we call f a facet of P*

Clearly, in the case of fully-dimensional polyhedra $dim(f) = n-1$ if f is a facet of P. If f is not a facet then there does not exist a unique representation for it. For example, consider a face f_o of a three-dimensional polyhedron P, defined as the line segment between two points in $\mathbb{R}^3$. Clearly, we need two inequalities to represent f_o since the line containing f_o is the intersection of two hyperplanes. Hence, the dimension of $f_o = n - 2 = 1$. In addition, there exists an infinite number of hyperplanes that intersect on the line containing f_o and any pair of these can represent f_o. Now consider a face f_1 of a three-dimensional polyhedron P that contains three affinely independent points. Clearly then, there exists a unique hyperplane containing f_1, i.e. f_1 has a unique support and there exists a unique inequality representing it.

A minimal representation of a polyhedron can be obtained by considering only the finite set of inequalities representing its facets, i.e.:

Definition 13 *Let $f_1, ..., f_t$ be the facets of the polyhedron $P = \{x \in \mathbb{R}^n | \mathbf{A}x \leq \mathbf{b}\}$, and let*

$(\mathbf{a}_1, b_1), ..., (\mathbf{a}_t, b_t)$ *define their supporting inequalities (halfspaces). Then:*

$$P = \{x \in \mathbb{R}^n | \mathbf{a}_i x \leq b_i, i = 1, ..., t\} \tag{2.8}$$

In other words, every polyhedron $P \subseteq \mathbb{R}^n$ can be represented (minimally) as the intersection of a finite number of halfspaces. We call this the representation of P in *interesection-form.*[4].

We now discuss an alternative way to represent polyhedra, which is dual to the representation we just gave. This representation is in terms of the *extreme points* and *extreme rays* of polyhedra, defined below:

Definition 14 *Consider the polyhedron* $P = \{x \in \mathbb{R}^n | \mathbf{A}x \leq \mathbf{b}\}$ *and its recession cone* $RC \subseteq \mathbb{R}^n$ *defined as* $RC = \{r \in \mathbb{R}^n | \mathbf{A}r \leq \mathbf{0}\}$. *Then:*

a) $x_o \in P$ *is an extreme point of* P *if and only if* $\nexists x_1, x_2 \in P : x_o = \lambda_1 x_1 + \lambda_2 x_2,\ \lambda_1 + \lambda_2 = 1.$

b) $r_o \in RC$ *is an extreme ray of* RC *and* P *if and only if* $\nexists r_1, r_2 \in RC$ *with* $r_1 \neq \alpha r_2,\ \alpha > 0 :$ $r_o = \lambda_1 r_1 + \lambda_2 r_2,\ \lambda_1 + \lambda_2 = 1\ \lambda_1, \lambda_2 \geq 0$.[5]

These notions are closely related to the notion of faces of polyhedral sets. More specifically, an extreme point of P is also a 0-dimensional face of P, i.e. the unique point of contact of a supporting hyperplane with P. Similarly, an extreme ray is a 1-dimensional face of RC. Now consider a polytope, i.e. a bounded polyhedron, $\Pi \subseteq \mathbb{R}^n$ and a polyhedral cone $C \subseteq \mathbb{R}^n$ and let $X = \{x_1, ..., x_{t1}\}$, $R = \{r_1, ...r_{t2}\}$ contain the t_1 extreme points and the t_2 extreme rays of Π and C respectively. Then:

$$\Pi = Conv(X) \tag{2.9}$$

$$C = Cone(R) \tag{2.10}$$

That is, a polytope can be expressed as the convex hull of its extreme points and a polyhedral cone as the conical hull of its extreme rays. Hence we shall say that any polytope and any polyhedral cone are *finitely generated.* We now generalise this representation to polyhedra. For

[4]Note that this concerns both fully-dimensional and non fully-dimensional polyhedra. In the latter case, there does necessarily exist a unique minimal representation. We do not discuss this case here.

[5]Note that the rays of C are unique up to scalar multiplication, i.e. the ray r_o is the same as αr_o where $\alpha > 0$.

any polyhedron $P \subseteq \mathbb{R}^n$, there exists a polytope $\Pi \subseteq \mathbb{R}^n$ and a polyhedral cone $C \subseteq \mathbb{R}^n$, such that:

$$P = \Pi + C \tag{2.11}$$

Note that + denotes set addition. In words, every polyhedron is the sum of a polytope and a cone. More specifically, for any $x_o \in P$, $\exists x' \in \Pi$ and $r' \in C : x_o = x' + r'$. This representation is sometimes referred to as the *Minkowski-sum* of P. Taking this further we can define a more precise representation for P, i.e. determine the polytope and the cone required for its description. As mentioned earlier this will be in terms of the extreme points and extreme rays of P:

Definition 15 *Consider the polyhedron* $P = \{x \in \mathbb{R}^n | \mathbf{A}x \leq \mathbf{b}\}$. *Let* $X^P = \{x_1, ..., x_{t1}\}$ *and* $R^P = \{r_1, ...r_{t_2}\}$ *contain its extreme points and extreme rays respectively. Then:*

$$P = Conv(X^P) + Cone(R^P) \tag{2.12}$$

It follows that every polyhedron is a finitely generated set. In the above $Conv(X^P)$ is a polytope and $Cone(R^P) = RC$. Finally, we close this section by briefly discussing the issue of *polarity.*

Definition 16 *Let* $P \subseteq \mathbb{R}^n$ *be a polyhedral set. The polar cone of* P, *is* $P^* = \{(y, y_0) \in R^{n+1} | \; yx \leq y_0 \; \forall x \in P\}$.

Note that in general the dimension of P^* is one more than the dimension of P. The polyhedral cone P^* contains any valid inequality for P. Let $(\mathbf{a}_o, b_o)$ define a facet supporting inequality for P. Then $(\mathbf{a}_o, b_o)$ is an extreme ray of P^*. In addition, if x_o and r_o are an extreme point and an extreme ray of P respectively, then $(x_o, -1)$ and $(r_o, 0)$ define facets of P^*, i.e. the hyperplanes defined by the equalities $yx_o - y_0 = 0$ and $yr_o = 0$ support facets of P^*. In general, there is a correspondence between the facets of P and the extreme rays of P^* and vice-versa.

Finally, note that for a polyhedral cone $C \subseteq \mathbb{R}^n$ any facet supporting inequality will be of the form $\mathbf{a}_o x \leq 0$. For this reason frequently in the literature its polar cone C^* is defined as $C^* = \{y \in \mathbb{R}^n | \; yx \leq y_0 \; \forall x \in C\}$. Using this definition C^* does not include all valid inequalities for C but just the supporting ones. In such a case $(C^*)^* = C$. For simplicity of presentation, we shall use this alternative definition when discussing the CRS DEA production possibility set in the following section.

2.3 Alternative representations of production possibility sets

2.3.1 The constant Returns-to-Scale case

Let us consider the production possibility set associated with the envelopment form of the CCR model given in (1.12) which we denote by T_C. It is easy to see that T_C is the collection of all feasible input-output bundles in $s+m-$dimensional space that satisfy the following properties:

1 Convexity.

2 Monotonicity.

3 Constant Returns to Scale (CRS), that is : $\forall(x,y) \in T_C \Rightarrow (\kappa x, \kappa y) \in T_C \; \forall \; \kappa \in \mathbb{R}_+$

4 Overall inclusion: $\forall j \in J \Rightarrow (\mathbf{x}_j, \mathbf{y}_j) \in T_C$

5 Minimum Extrapolation: T_C is the closed intersection of all possible sets that satisfy the above properties.

We now explore two alternative ways to represent T_C. The two representations in fact form a dual pair. The first, also referred to as primal representation, is in terms of non-negative linear combinations of observed bundles (i.e. DMUs):

$$T_C \equiv \{(x,y) \in \mathbb{R}_+^{s+m} |\; x \geq \mathbf{X}\boldsymbol{\lambda},\; y \leq \mathbf{Y}\boldsymbol{\lambda},\; \boldsymbol{\lambda} \geq 0\} \tag{2.13}$$

It is easy to see that for any $(x,y) \in T_C$ there exists a feasible solution for model (1.12). We now examine how to obtain a minimal primal representation of T_C. Let $\hat{E}_C \subseteq J$ be the set of extreme-efficient DMUs for the CCR model[6], and Z_C be the set of their data points, i.e. $Z_C = \{(\mathbf{x}_j, \mathbf{y}_j) \in \mathbb{R}_+^{s+m} |\; \forall j \in \hat{E}_C\}$. We also define the set of *free disposability rays* $R = \{r_k \in \mathbb{R}^{s+m} |\; k \in (1,...,s+m),\; k \leq s \Rightarrow r_k = e_k,\; k > s \Rightarrow r_k = -e_k \}$[7]. Below we obtain the minimal primal representation of T_C:

[6] A CCR Pareto-efficient DMU is extreme-efficient if it cannot be expressed as a non-neggative linear combination of other DMUs (see also Charnes et al. (1991)).

[7] e_k is a vector with 1 in the k'th position and all other entries equal to zero.

Lemma 1 *Let T_C be defined as in (2.13). Then, T_C can be represented minimally as follows:*

$$T_C = Cone(Z_C \cup R) \cap \mathbb{R}_+^{s+m} \tag{2.14}$$

Proof. To establish that $Cone(Z_C \cup R) \cap \mathbb{R}_+^{s+m} \subseteq T_C$, consider a point $(x, y) \in Cone(Z_C \cup R) \cap \mathbb{R}_+^{s+m}$. If $(x_o, y_o) \in Cone(Z_C) \cap \mathbb{R}_+^{s+m}$ (which is contained in $Cone(Z_C \cup R) \cap \mathbb{R}_+^{s+m}$), then $(x_o, y_o) \in T_C$. Otherwise if (x_o, y_o) is a nonnegative combination of some DMUs in $\hat{E}_C$ and some rays in R, then clearly there exists $(x', y') \in Cone(Z_C) \cap \mathbb{R}_+^{s+m}$ such that $x' \leq x_o$ and $y' \geq y_o$. But $(x', y') \in T_C$, hence $(x_o, y_o) \in T_C$. An analogous argument can be made to establish that $T_C \subseteq Cone(Z_C \cup R) \cap \mathbb{R}_+^{s+m}$. ■

To obtain the dual representation we utilise P_C, the polyhedral cone of input-output weights (*weighting schemes*) corresponding to T_C. This is the polar cone of T_C and it contains all feasible solutions to (1.11) without the normalisation constraint $vx_o = 1$, i.e.:

$$P_C \equiv \{(v, u) \in \mathbb{R}_+^{s+m} |\ -v\mathbf{x}_j + u\mathbf{y}_j \leq 0, \forall\, j \in J\}^8 \tag{2.15}$$

The following theorem (see Olesen and Petersen 1996) shows that we can obtain an alternative representation of T_C in terms of P_C.

Theorem 1 *Let the primal representation of T_C be defined as in (2.13). Then, the dual representation of T_C is:*

$$T_C = \{(x, y) \in \mathbb{R}_+^{s+m} | \forall (v, u) \in P_C : -vx + uy \leq 0\} \tag{2.16}$$

Noting that P_C is a polyhedral cone and hence finitely generated, we obtain the following:

Corollary 1 *Let $\bar{P}_C$ be the set of extreme rays of P_C, and I_C its corresponding index set. Then,*

$$T_C = \{(x, y) \in \left(\bigcap_{i \in I_C} -v_i x + u_i y \leq 0 \right) \cap \mathbb{R}_+^{s+m} |\ (v_i, u_i) \in \bar{P}_C\} \tag{2.17}$$

[8] Strictly speaking, if inputs are represented positively, this should be $P_C \equiv \{(-v, u) \in \mathbb{R}^{s+m} |\ -v\mathbf{x}_j + u\mathbf{y}_j \leq 0, \forall\, j \in J\}$. We use this slightly different notation throughout this text for the sake of simplicity. Note also that as we mentioned earlier, we have defined P_C to contain only the supporting inequalities for T_C, not all valid ones. For this reason both T_C, $P_C \subseteq R^{s+m}$.

As discussed in the previous section, any polyhedral set can be represented either in *sum-form* with use of its extreme points and/or extreme rays (polyhedral cones have only one extreme point: the origin), or in *intersection-form* by using its facets. In the case of T_C, the sum-form representation is given by the conical hull of the extreme rays as in (2.14). The facet representation is given by the intersection of a collection of halfspaces as in (2.17). Since T_C is a finitely generated set, only a finite collection of halfspaces is needed to describe it as given in (2.17). In fact, as can be seen from the above, there is a correspondence between the facets of T_C and the extreme rays of P_C. More specifically, each extreme weighting scheme (ray) of P_C defines a hyperplane in the form $-vx + uy = 0$ (see also Ali and Seiford 1993) which contains a facet of T_C.

2.3.2 The variable Returns-to-Scale case

Now consider the BCC model in envelopment form, given in (1.20), and denote its production possibility set by T_V. In parallel to the CRS case, the primal representation of T_V is as follows:

$$T_V \equiv \{(x,y) \in \mathbb{R}_+^{s+m} |\ x \geq \mathbf{X}\boldsymbol{\lambda},\ y \leq \mathbf{Y}\boldsymbol{\lambda},\ \boldsymbol{\lambda} \geq 0,\ e\boldsymbol{\lambda} = 1\} \tag{2.18}$$

Let $\hat{E}_V \subseteq J$ be the set of extreme-efficient DMUs for the BCC model[9] and Z_V contain their data points. Similar to T_C, we obtain the minimal primal representation of T_V as the *Minkowski-sum* of the convex hull of its extreme points and the conical hull generated by the free disposability rays, i.e.:

Lemma 2 *Let T_V be defined as in (2.18). Then, T_V can be represented minimally as follows:*

$$T_V = (Conv(Z_V) + Cone(R)) \cap \mathbb{R}_+^{s+m} \tag{2.19}$$

Proof. Directly analogous to the proof for lemma 1 ■

The polar cone to T_V, denoted P_V is as follows:

$$P_V \equiv \{(v,u,\beta) \in \mathbb{R}_+^{s+m+1} |\ -v\mathbf{x}_j + u\mathbf{y}_j \leq \beta, \forall\ j \in J\} \tag{2.20}$$

[9] A BCC Pareto-efficient DMU is extreme efficient if it cannot be expressed as convex combination of any other DMU.

where β corresponds to the additional variable in multiplier VRS models. As before, we obtain a dual representation of T_V in terms of P_V.

Theorem 2 *Let the primal representation of T_V be defined as in (2.18). Then, the dual representation of T_V is:*

$$T_V = \{(x,y) \in \mathbb{R}_+^{s+m} | \forall (v,u,\beta) \in P_V : -vx + uy \leq \beta\} \tag{2.21}$$

Corollary 2 *Let $\bar{P}_V$ be the set of extreme rays of P_V, and I_V its corresponding index set. Then,*

$$T_V = \{(x,y) \in \left(\bigcap_{i \in I_V} -v_i x + u_i y \leq \beta_i \right) \cap \mathbb{R}_+^{s+m} | \ (v_i, u_i, \beta_i) \in \bar{P}_V\} \tag{2.22}$$

In the VRS case the supporting hyperplanes are defined by extreme rays of P_V as $-vx + uy = \beta$[10]. Finally, note that since $\alpha(-vx + uy) \leq \alpha\beta$ for any positive α, all positive multiples of a particular weighting scheme define the same halfspace/hyperplane.

Going back to the DEA models discussed in the previous chapter, suppose DMU k lies on a particular facet-defining hyperplane associated with weighting scheme $(v^*, u^*, \beta^*) \in \bar{P}_V$ and assume further that $v^* x_k \neq 0$. Then $u^* y_k - v^* x_k = \beta^* \Rightarrow u^* y_k / v^* x_k - \beta^* / v^* x_k = 1$ and also $u^* y_j / v^* x_k - v^* x_j / v^* x_k \leq \beta^* / v^* x_k \ \forall j \in J$. Hence we obtain that $(v^*/v^*x_k,\ u^*/v^*x_k,\ \beta^*/v^*x_k)$ is an optimal solution to (1.21) with corresponding optimal objective function value equal to one, i.e. DMU k is radially-efficient with $(v^*/v^*x_k,\ u^*/v^*x_k,\ \beta^*/v^*x_k)$ as an optimal weighting scheme.

2.3.3 Examples

Figure 2.1 illustrates the halfspace representations of the CRS and VRS production possibility sets for the example used earlier in figure 1.6. Six halfspaces are shown, each associated with one of the six hyperplanes h_1 to h_6, defined by a unique weighting scheme. The CRS production possibility set is the intersection of the halfspace associated with hyperplane h_3 with the non-negative orthant. The ray from O through B and C (contained in h_3) serves as the CRS efficient frontier (shown in red). The VRS production set is the intersection of all six halfspaces with the non-negative orthant, and the VRS efficient frontier is given by ABCDE[11] (shown in blue).

[10] In the CRS case β is always equal to zero. β is also referred to as the *intercept* or the *offset*.
[11] This is the Pareto-efficient part of the frontier

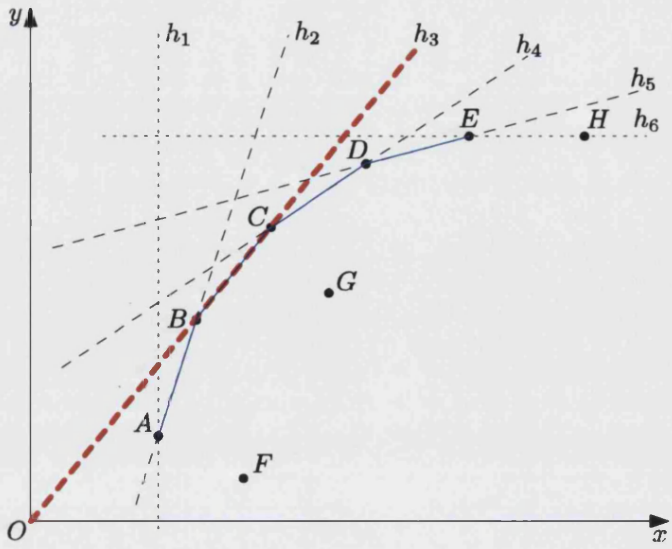
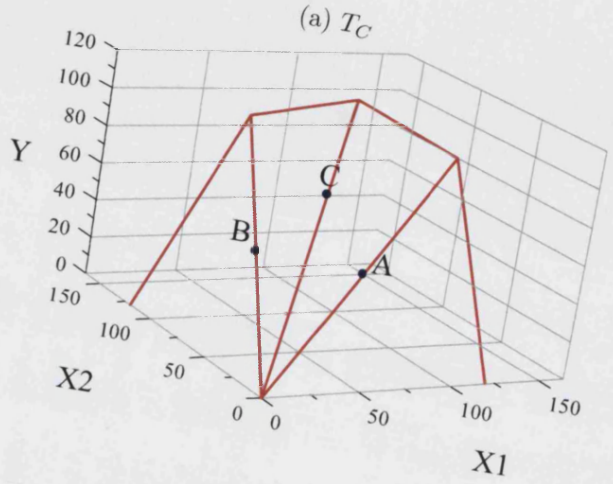


Figure 2.1: Halfspace representations of production possibility sets in two dimensions

To illustrate better the relative shape of the CRS and VRS production sets, as well as their dual representations, we use the three-dimensional examples in figures 2.2, 2.3 and 2.4, with three DMUs using two inputs $X1$ and $X2$ to produce one output Y. Figure 2.2a shows the primal representation of the generated CRS production possibility set, which is a polyhedral cone from the origin with extreme rays passing through points A-C. Figure 2.2b shows the primal representation of the generated VRS production possibility set, which is a polyhedral region of which the surface ABC is a facet. Note that the VRS polyhedral set is entirely contained within the CRS polyhedral cone as illustrated in figure 2.3. There are four facets in the CRS case and seven in the VRS case. We plot the halfspace representation of the CRS production set only, in figure 2.4b. The hyperplanes needed to describe the necessary halfspaces are equal in number to the number of facets for T_C, so that each of these hyperplanes supports one facet. All facets of T_C contain the origin so that all of the associated hyperplanes penetrate the origin and hence can all be described in general as $-vx + uy = 0$. In the VRS case the associated hyperplanes are not guaranteed to contain the origin, so the additional term β is needed for their description. Finally, note that in both cases all DMUs lie on more than one facet/hyperplane which implies that they possess more than one set of optimal weights (we discuss this issue further in subsequent chapters).



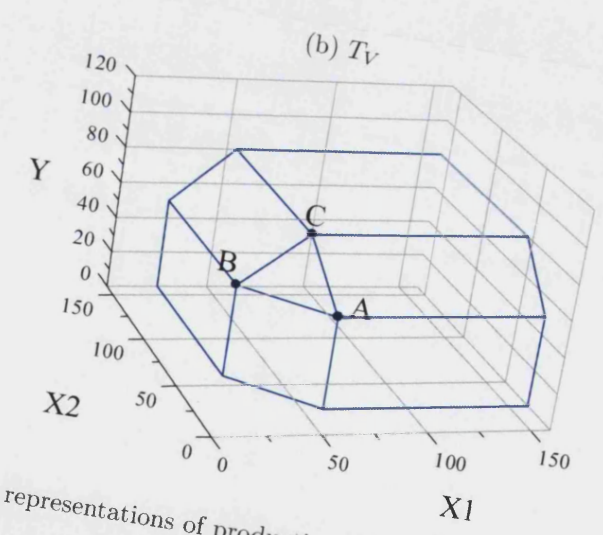


Figure 2.2: Primal representations of production possibility sets in three dimensions

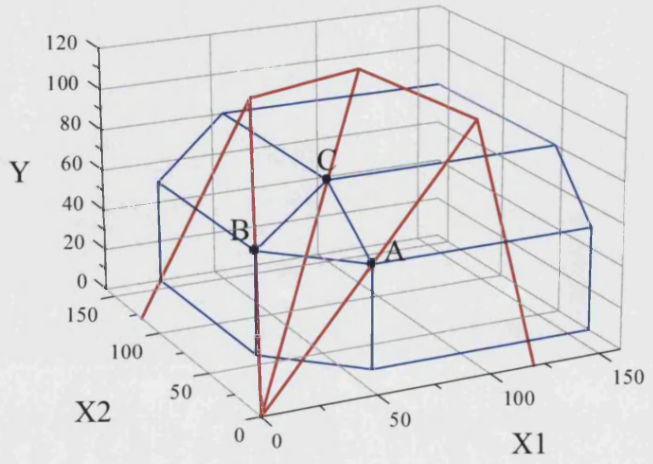


Figure 2.3: T_V contained in T_C

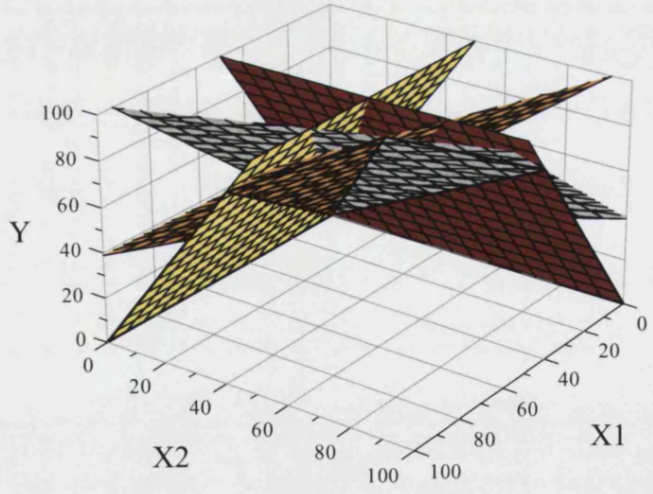


Figure 2.4: Dual representation of T_C in three dimensions

2.4 Identification of production possibility sets

This section reviews the existing approaches reported in the DEA literature for explicitly identifying the DEA production possibility sets. By this we mean the explicit identification, or computation, of either the complete set of supporting hyperplanes, or, equivalently, all subsets of extreme-efficient DMUs, along with the relevant extreme rays, that span all of its facets. We categorise these approaches in two types, namely optimisation approaches which involve the use of mixed integer and linear programming models, and convex hull approaches which utilise algorithms for the identification of the convex hull of a set of points. In the subsequent discussion we shall assume that the explicit identification of the production set is the second part of a two-stage process, where in the first stage the set of all extreme-efficient DMUs has been identified ($\hat{E} \subseteq J$). Non-extreme DMUs are not necessary in generating any faces of the DEA production set and hence including them in the second stage would only increase the difficulty of what is already a very hard problem.

2.4.1 Optimisation approaches

Approaches using Mixed Integer Programmes (MIPs) for identifying all facets of a DEA production set were the first to be suggested in the literature. A predecessor to these approaches is the model in Olesen and Petersen (1996) used for establishing the existence of fully-dimensional faces of a production possibility set which contain Pareto-efficient production bundles only. In the spirit of this model, all subsequent MIP based approaches focused on the identification of fully-dimensional efficient faces only. We shall give a general description of these approaches.

Following Raty (2002), we can define the set of all feasible weighting schemes that define fully-

dimensional efficient faces in a VRS technology as follows:

$$
\begin{aligned}
P_F = \{ \quad & u\mathbf{Y} - v\mathbf{X} - e\beta + \mathbf{s} = \mathbf{0} \\
& \mathbf{s} \leq (1 - \mathbf{b})\mathcal{M} \\
& e\mathbf{b} \geq s + m \\
& v_i = 1 \text{ for some } i \in \{1, ..., s\} \\
& b_j \in \{0, 1\} \; \forall j \in \hat{E}_V \\
& (v, u) \in \mathbb{R}_+^{s+m}, \; \beta \in \mathbb{R}, \; \mathbf{s} \in \mathbb{R}_+^{|\hat{E}_V|} \\
& \mathcal{M} \text{ is a very large positive constant}\}
\end{aligned}
\tag{2.23}
$$

where the matrices $\mathbf{X}$ and $\mathbf{Y}$ only contain the input-output data for the set $\hat{E}_C$ of extreme-efficient DMUs. Note that the two constraints $\mathbf{s} \leq (1 - \mathbf{b})\mathcal{M}$ and $e\mathbf{b} = s + m$, together force at least $s + m$ components of $\mathbf{s}$ to be zero and hence a feasible solution to (2.23) will identify a face that contains at least $s+m$ extreme-efficient DMUs. In other words the set of feasible solutions to (2.23) contains all hyperplanes that support at least $s + m$ extreme-efficient DMUs.

However, as we will demonstrate in the next chapter, such hyperplanes do not necessarily support facets that contain Pareto-efficient production bundles only. As a result, approaches based on the above set of constraints will end up identifying additional facets that contain weakly-efficient production bundles. Nevertheless, the results in the next chapter will also show that we can always identify and exclude such facets posterior to the analysis by inspecting their defining weighting schemes.

An iterative optimisation procedure can be constructed by including the set of constraints in (2.23) (in some form) along with some other constraints to guarantee that a new efficient face is identified in every iteration. The procedure can be either applied to all extreme-efficient DMUs separately, or it can involve a global MIP solved as many times as necessary for the identification of all relevant facets. An approach of the latter type is given by Raty (2002), who reports the use of a 'bogus' (constant) objective function in a code that can identify all facets for a 'moderate' number of extreme-efficient units. Olesen and Petersen (1996) develop an MIP that is applied to each DMU separately along the same lines. Their model is based on the consecutive introduction

of cutting-planes to avoid the identification of faces already found. For every DMU, the solution process is repeated until infeasibility is encountered, i.e. no additional facets spanned by that DMU can be identified, at which point the loop proceeds to the next DMU. Very similar models are also reported by Jahanshahloo et al. (2005) and Amirteimoori and Kordrostami (2005).

Another set of optimisation approaches in the DEA literature is based on the solution of LPs to identify all facets of DEA production sets. From solving any multiplier DEA model for a particular DMU, we obtain a set of weights which define a hyperplane supporting a facet of the production set. Hence, to identify all facets this way we would essentially need to identify all possible basic solutions to the multiplier DEA LPs for all DMU instances, which is what these approaches work towards. Huang et al. (1993, 1997) generate new facets by solving an additional multiplier DEA LP for each efficient DMU and identify which of its peers satisfy their constraints as equalities, i.e. the subset of DMUs that lie on the same efficient facet as that DMU. Their claim is that by applying these to all DMUs various subsets are identified " and there will be duplication, overlapping and 'nesting' of these various subsets from which ... [facets] can be identified by reduction". However, the presence of degeneracy/multiple-optimality in many instances of DEA models does not allow for the exhaustive identification of facets by this approach. This is also noted by Pitakong et al. (1998) who developed modified pivoting criteria for the *simplex method* as alternative strategies for identifying more efficient facets. Although this approach claims to identify more facets, it offers no guarantee that all efficient facets can be identified (as noted by the authors).

An exhaustive LP approach is introduced by Yu et al. (1996b). This is based on the *Generalised DEA model*, also introduced by the same authors (Yu et al. 1996a), and the use of *predilection cones* to identify all facets. Predilection with regards to a pre-specified subset of DMUs refers to a set of additional constraints introduced in a DEA model in order to identify DMUs that are on the same facet as those in the pre-specified subset. This gives rise to an iterative procedure equivalent to an enumerative tree search, similar to the procedure for identifying all cliques of an undirected graph by Bron and Kerbosch (1973), during which at some stage a subset of DMUs already found to be on the same facet "is augmented by the addition of another suitably chosen DMU [also found to be on the same facet] ... until no further augmentation is possible and the subset becomes a maximal subset". Unfortunately, no computational experience for this interesting approach is reported in the DEA literature.

2.4.2 Convex Hull approaches

The second set of approaches implement convex hull algorithms for explicitly identifying the DEA production set in terms of facets. Clearly, the convex hull of a set of given points is a polytope and hence can be represented either in sum-form by its extreme points (in terms of convex combinations) or in intersection-form as the intersection of a finite collection of closed halfspaces. These two problems are respectively referred to as the *vertex* and *facet enumeration problems.* Moving from one representation to the other, which is precisely what we need to do in the DEA case, is a notoriously hard problem and so far all proposed algorithms are of exponential complexity. Nevertheless, the advantages of providing a complete representation of the production sets make the additional computational effort worthwhile in datasets for which this can be achieved in reasonable time.

To be more precise, the above description of the convex hull problem relates to a polytope, i.e. a bounded polyhedron whereas the DEA production sets are unbounded polyhedra for which the sum-form representation additionally requires the identification of their extreme rays. Especially for the CRS case, the production possibility set is a polyhedral cone and hence only has one extreme point (the origin) and a subset of DMUs as extreme rays (plus some additional extreme rays). The theory in the previous section tells us that we can identify all facets of a DEA production set by feeding into the implementation algorithms the corresponding sets of extreme efficient DMUs and extreme rays. In the CRS case the origin is specified as the unique extreme point and $Z_C \cup R$ is used as the set of extreme rays, whereas in the VRS case Z_V and R contain the extreme points and rays respectively.

Many codes exist that can identify unbounded polyhedra explicitly in terms of facets (given the extreme points and rays). The use of two such codes is reported in the approaches by Appa and Williams (2006) and Briec and Leleu (2003). These respectively use *PORTA*[12] and *cdd*[13]. The implemented algorithms are closely related. PORTA is based on the *Fourier-Motzkin Elimination* (FMEL) method for solving a system of linear inequalities (see Williams 1986; Dantzig and Eaves 1973), and *cdd* is based on the *Double Description Method* of Motzkin et al. (1953), a variant of

[12] Polyhedron Representation Transformation Algorithm, a code by Thomas Christof, Heidelberg University, and Andreas Loebel, Konrad-Zuse-Zentrum fur Infromatik, Berlin (ZIB)

[13] Code by Komei Fukuda, Swiss Federal Institute of Technology

FMEL.

Appa and Williams (2006) introduce the application of FMEL to the dual side of the problem in order to identify directly the extreme rays of P_C and P_V. More specifically, they apply FMEL to the set of constraints in (1.11) or (1.21) which are common to all DEA model instances (all DMUs' models) and identify all possible optimal weighing schemes, which they call the *complete set of weights.* This directly provides all necessary supporting hyperplanes. Their approach is also utilised in Appa, Argyris, and Williams (2006) who report the computability of problems with $|\hat{E}_C| \leq 50$ and $s + m \leq 10$ in reasonable time.

Olesen and Petersen (2003) additionally report the use of *Qhull*[14] for the identification of all efficient facets of the production set (as well as the previous codes). *Qhull* uses floating point arithmetic and hence can be much faster than *cdd* and *PORTA* albeit less precise. *Qhull* does not directly support the envelopment of both extreme points and rays and hence some manipulation of the dataset is needed. Olesen and Petersen's approach utilises points at infinity (by employing big Ms), although this introduces some extra theoretical concerns. Unlike in Appa and Williams's approach, many of the identified facets do not necessarily correspond to the DEA production set and further inspection is needed to select those that do. In addition, for the CRS case Olesen and Petersen's *Qhull* approach cannot directly identify the conical-hull of data points. Instead it computes the convex hull of all data points plus the origin and then discards the ones with non-zero offset term β. This results in considerable additional computational effort[15]. Nevertheless, the use of *Qhull* allows much bigger problems to be solved much faster. Olesen and Petersen report the explicit identification of T_C for a problem with $|\hat{E}_C| = 136$ and $s + m = 13$ in reasonable time. With respect to the computability of larger datasets, Olesen and Petersen estimate upper bounds for the number of extreme-efficient DMUs and the number of inputs and outputs, at $|\hat{E}_V| \leq 1000$ and $s + m \leq 25$ respectively.

[14] Code by Brad Barber, David Dobkin and Hannu Hundanpaa, Universitu of Minesota

[15] Compared to the computations that would have been required had *Qhull* been able to support envelopment of both extreme points and rays. Compared to other codes, *Qhull* should be faster in any case because of the use of floating point arithmetic.

2.4.3 Examples

To illustrate how convex hull approaches identify all facets of a DEA production set we include a small two-dimensional example of five DMUs given in table 2.1. We consider the CRS case only. Because this is a very small example there is no need to identify all extreme-efficient DMUs in the first phase. Let the set of data points be $Z = \{(3,3)^T, (6,5)^T, (9,6)^T, (13,7)^T, (8,4)^T\}$[16] and the set of free-disposability rays be $R = \{r_1 = (1,0)^T,\ r_2 = (0,-1)^T\}$. An illustration is given in figure 2.5. Figure 2.5a presents the data points and the free disposability rays. To obtain the dual description of T_C we feed the set of rays $Z \cup R$ into *PORTA* or *cdd*. The output obtained is a minimal set of inequalities needed to represent $Cone(Z \cup R)$. This includes two inequalities, namely $-x + y \leq 0$ and $-x \leq 0$ given in figure 2.5b. Since $T_C = Cone(Z \cup R) \cap \mathbb{R}_+^{s+m}$ the later inequality is dropped, as only the first one is needed[17] to represent T_C. Figure 2.5c illustrates Appa and Williams' approach. We work directly in the dual space by providing inequalities and obtaining the extreme rays of P_C. In particular, the set of inequalities $(-v, u)Z \leq 0,\ v \geq 0,\ v \geq 0$ is fed into PORTA, these are the five inequalities denoted $a - e$ along with inequalities x and y which represent the nonnegativities $(u, v \geq \mathbf{0})$. Two extreme rays for $Cone(Z \cup R)$ are identified, namely $(1,1)$ and $(1,0)$ which clearly correspond to the inequalities described by the previous approach. In the same fashion, the second ray is dropped so that the unique extreme ray $(1,1)$ of P_C is identified as relevant for the explicit characterisation of T_C. Finally, in figure 2.5d we illustrate Olesen and Petersen's approach. The set of data points Z along with the origin and two points at infinity are fed to *Qhull*. Their convex hull is then identified in terms of six inequalities. From these we only keep those that are defined by hyperplanes that include the origin. There is only one such hyperplane (f_2) and, as before, this is the only one required for a dual representation of T_C.

[16] Note that we can also represent the inputs negatively in oder to directly obtain positive weights for all inputs. For simplicity of presentation we overlook this aspect and assume that all weights obtained are positive.

[17] To be precise, and as was also noted earlier, the inequality $y \geq 0$ is also needed. Note that the inequalities $x \geq 0$, $y \geq 0$ are common for every CRS production set. Furthermore, it does not define a set of valid DEA weights. Hence, it is disregarded. In general we are only interested in inequalities with neggative input (if inputs are represented positively) and positive output multipliers that are of $\leq$ direction.

DMU	Input (x)	Output (y)
A	3	3
B	6	5
C	9	6
D	13	7
E	8	4

Table 2.1: Example dataset

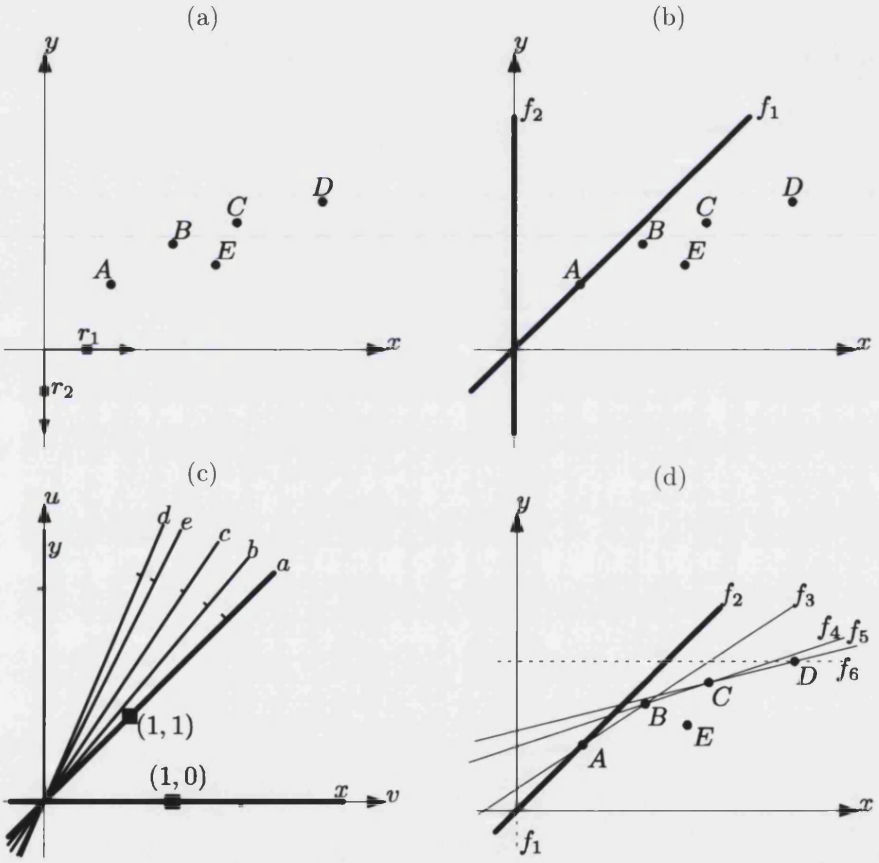


Figure 2.5: Identiying the production possibility set explicitly

Chapter 3

On facets, dimension and efficiency

3.1 Introduction

In this chapter we explore further the attributes of efficient production in terms of DEA production possibility sets. We contribute to the existing research in the field by developing a set of theoretical constructs that associates the concepts of efficiency and weak-efficiency with specific segments of the boundary of the DEA production possibility sets as well as their polar cones.

The relevant discussion in the DEA-related literature has explored this issue mostly on a 'micro-level', by examining and imposing conditions on the optimal solutions of the various DEA models. In particular, the two main issues this discussion revolves around are: a) the use of zero weights by DMUs and b) the presence of slacks at their optimal targets (see e.g. Charnes et al. 1991; Thrall 1996; Portela and Thanassoulis 2006).

More recently, the discussion about efficiency related to specific characteristics of production possibility sets has appeared in the literature (Olesen and Petersen 1996, 2003), and has introduced additional concerns regarding dimensional aspects of segments of the efficient frontier and the issue of *well-defined rates of substitution.*

However, the existing theoretical framework for this is incomplete. Our new developments address this issue; they systematise existing research, address inconsistencies that we discovered and, most importantly, fill many 'gaps' by introducing new theoretical constructs.

In particular, we develop necessary and sufficient conditions for the existence of fully-dimensional segments of the efficient frontier and *well-defined rates of substitution.* These results allow for complete characterisation of faces of the production possibility sets in terms of their efficiency classifications. More specifically, our results show how to characterise efficient and weakly-efficient surfaces of the efficient frontier, i.e. surfaces that only contain Pareto-efficient input-output bundles and surfaces that also contain weakly-efficient bundles. In addition, the new framework allows us to critically examine the existing research in the field. This leads us to discover important inconsistencies which we also demonstrate by providing counter-examples.

The outline of this chapter is as follows. We start by providing formal definitions relating to the efficiency classification and dimension of faces of the production possibility set in section 3.2. Building on these definitions we develop theoretical constructs for these concepts in section 3.3, by considering them separately at first and subsequently combining the two separate sets of results. Using our new results we are able to detect inconsistencies in the related literature. We do this in section 3.4, where we address two specific problems. These concern the number of extreme-efficient DMUs contained in faces of the production possibility set and the weighting schemes that define such faces. For both cases we discuss the problems by relating them to our new results and provide counter-examples for the existing results in the literature.

3.2 Efficiency and dimension of faces: Definitions

In this section we provide proper definitions for characterising the efficiency classification of facets as well as their dimension. Our discussion focuses on efficient production, hence we shall concentrate on input-output bundles that lie on the boundary of the production possibility sets. The results extend, of course, to inefficient (interior) bundles which are radially projected on the frontier by DEA models. Note that by production bundles we do not necessarily mean DMUs, but input-output vectors included in the production possibility set in general.

To aid with the theoretical discussion we will be using an example of nine DMUs given in table 3.1. The VRS production possibility set generated by these units is given in figure 3.1. Firstly, note that all units are extreme-efficient. The boundary of the production set is comprised of eighteen facets. We use A′, B′ etc. to indicate points at infinity, so that C′CAA′, for example,

DMU	Input 1 ($X1$)	Input 2 ($X2$)	Output (Y)
A	60	25	50
B	25	60	50
C	45	45	80
D	75	75	120
E	125	95	135
F	95	125	135
G	205	125	150
H	125	205	150
I	180	180	165

Table 3.1: Example dataset

is an 'unbounded' facet. We refer to such facets as weakly-efficient facets because they clearly contain weakly-efficient input-output bundles and hence are included in the Pareto-efficient part of the DEA frontier. We use the term efficient facets to refer to facets that only contain Pareto-efficient bundles. Table 3.2 gives the weighting schemes that define the supporting hyperplanes of the production set[1]. Recall that the offset β corresponds to additional variable associated with VRS multiplier models. For convenience, in figure 3.2 we provide a *graph* of the *facial structure* of T_V. The *nodes* (or *vertices*) A-I of the graph correspond to extreme points of polyhedron T_V (extreme-efficient DMUs) and the nodes r_1, r_2 and r_3 correspond to the free disposability rays relating to the two inputs and the output respectively (extreme rays). Two nodes are linked by an *edge* (or *link*), only if together they are necessary for the description of a face of T_V. For example nodes CD are linked because they define a face of T_V and nodes C and r_1 because they describe the face CC′ of T_V. Any maximal subset of nodes in the graph such that any pair of nodes in this subset is connected by an edge[2], defines a facet of T_V. We call such subsets of nodes *cliques* of the graph.

Olesen and Petersen (1996, 2003) introduced the term *efficient facets* to refer to efficient faces of the production set that are maximal under inclusion, i.e. not contained in other efficient faces. For example in figure 3.1, ABC and CD are both efficient facets. Clearly then, this definition implies that an efficient facet is not necessarily a fully-dimensional face of the polyhedral production

[1] Strictly, speaking this production set has an additional facet, namely the one contained in hyperplane $y = 0$, implied by the nonnegativity of output. In general, for every production possibility set, all nonnegativities ($x_i \geq 0$ $i = (1, ..., s)$, $y_r \geq 0$ $r = (1, ..., m)$) can possibly define facets. Note that these facets may contain inefficient bundles. To maintain simplicity, and also because such facets would be common to all DEA production sets, we shall completely disregard them from our analysis.

[2] A subset S of pairwise-connected nodes is maximal if there does not exist another node in the graph which is connected to every node in S.

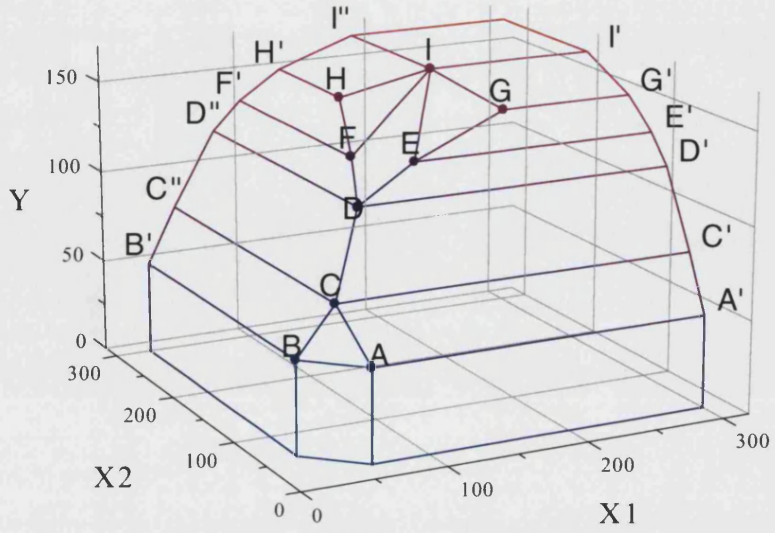


Figure 3.1: The VRS production possibility set generated the DMUs in table 3.1

Supporting Hyperplane	Associated Weighting Scheme $(v_1,\ v_2,\ u,\ \beta)$	Spanning DMUs
H_1	(6, 6, 1, −460)	(A, B, C)
H_2	(1, 1, 0, −85)	(A, B)
H_3	(1, 0, 0, −25)	(B)
H_4	(0, 1, 0, −25)	(A)
H_5	(3, 0, 2, 25)	(B, C)
H_6	(0, 3, 2, 25)	(A, C)
H_7	(4, 0, 3, 60)	(C, D)
H_8	(0, 4, 3, 60)	(C, D)
H_9	(3, 0, 4, 255)	(D, F)
H_{10}	(0, 3, 4, 255)	(D, E)
H_{11}	(3/950, 1/1130, 103/9975)	(F, H, I)
H_{12}	(1/1130, 3/950, 103/9975)	(E, G, I)
H_{13}	(3, 0, 11, 1275)	(H, I)
H_{14}	(0, 3, 11, 1275)	(G, I)
H_{15}	(1, 0, 2, 175)	(F, H)
H_{16}	(0, 1, 2, 175)	(E, G)
H_{17}	(3, 3, 14, 1230)	(D, E, F, I)
H_{18}	(0, 0, 1, 165)	(I)

Table 3.2: Hyperplanes and weighting schemes for the DMUs in table 3.1

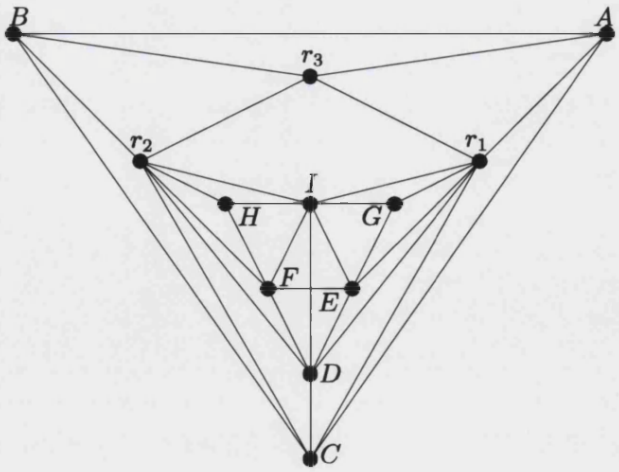


Figure 3.2: Graph of the facial structure of T_V

possibility set, and hence departs from the standard definition of facets in convex analysis and polyhedral theory (definition 12). To avoid confusion, we introduce a slightly different terminology and refer to maximal efficient faces as *efficient surfaces*. In contrast we use the term weakly-efficient surfaces to refer to maximal weakly-efficient faces.

In what follows we will examine in detail the twin concepts of face dimension and face efficiency classification. The following two definitions related to these concepts will provide the basis for our discussion.

Definition 17 *Let H_a be a supporting hyperplane for T_C and H_b a supporting hyperplane for T_V and consider the faces f_a and f_b of T_C and T_V respectively, defined as $f_a = H \cap T_C$ and $f_a = H \cap T_V$. Then, f_a (or f_b) is:*

(a) An efficient face, if and only if all production bundles contained in it are Pareto-efficient.

(b) A weakly-efficient face, if and only if it contains at least one weakly-efficient production bundle.

Looking back in figure 3.1, faces CD and ABC are efficient faces and faces AA′ and B′BCC″ are weakly-efficient faces. Note that CD is not a fully-dimensional efficient face of T_V. With regards to face dimensionality, we use the following definition:

Definition 18 *A face f of T_C (or T_V) is fully-dimensional (i.e. a facet) if and only if it contains*

$s+m$ affinely independent production bundles.

Here we have departed from definitions appearing in the related literature. Considering the discussion in the previous chapter, the rationale behind our definition is that a face is fully-dimensional if and only if it contains a sufficient number of vectors such that their affine hull equates to the supporting hyperplane to the particular face. Of course, some faces may be contained in many hyperplanes, but as we shall see below this is not the case for fully-dimensional faces (facets). For example, in figure 2.2a, face ABC is a fully-dimensional face of T_V as it clearly contains $s+m=3$ affinely independent production bundles, whereas face AB contains no more than two and hence is not fully-dimensional. With respect to the CRS case, it is important to note that the origin can be included in these $s+m$ vectors. For example, in the same figure the face of T_C spanned by the origin and DMUs A and B is fully-dimensional, because these three vectors are affinely independent, whereas face OA is not. We formalise these observations in the following remark. It is important to note that this *only* provides *sufficient* conditions.

Remark 1 *Let f_a and f_b be faces of T_C and T_V respectively. Then:*

(a) If f_a contains at least $s+m-1$ extreme-efficient units, then it is a fully-dimensional face of T_C.

(b) If f_b contains at least $s+m$ extreme-efficient units, then it is a fully-dimensional face of T_V.

Definitions appearing in the related literature use the concept of linear instead of affine independence. For example, Portela and Thanassoulis 2006 define fully-dimensional faces as those containing $s+m-1$ linearly independent Pareto-efficient DMUs in the CRS case and $s+m$ in the VRS case.

We would like to point out that the use of linear independence however is *problematic* in the VRS case. Consider the production set T_V in figure 2.1. According to their definition face BC contained in hyperplane f_3, which is clearly a facet, is not fully-dimensional since C can be expressed as a multiple of B (and vice versa). In general, this problem concerns faces that are supported by hyperplanes which penetrate the origin. Such hyperplanes are subspaces of $\mathbb{R}^n$ and cannot contain more than $s+m-1$ linearly independent points, irrespective of whether they support faces of a CRS or a VRS production set. Hence, imposing different requirements in assessing the dimension

of faces based on the Returns-to-Scale specification creates this problem. Our definition can clearly work around this. The problem does not arise in the CRS case because all supporting hyperplanes of T_C penetrate the origin so the requirement of the existence of $s + m - 1$ linearly independent Pareto-efficient DMUs is justified. In such a case, the same set of DMUs plus the origin define a set of $s + m$ affinely independent vectors so that we would be able to identify the corresponding face as fully-dimensional using our definition as well.

3.3 Theoretical developments

Having provided proper definitions, we now examine the concepts of face-efficiency and face-dimension in more detail. We start by considering these separately and establishing a set of theoretical results for each. Subsequently, we combine our results to establish a classification scheme for all faces with respect to both of these concepts.

3.3.1 Face-efficiency

Going back to our example, we can see in figures 3.1 and 3.2 that some facets can be expressed by a convex combinations of extreme points while some others, the unbounded ones, are expressed by taking combinations of some extreme points and rays[3]. This connection is formalised below.

Lemma 3 *Let f_a be a face f of T_C and f_b a face of T_V. Then f_a (or f_b) is:*

(a) An efficient face, if and only if only extreme-efficient DMUs (with data points in Z_C or Z_V) are needed for its primal description.

(b) A weakly-efficient face, if and only if both extreme-efficient DMUs (in Z_C or Z_V) and free-disposability rays (in R) are required for its primal description.

Proof. One only needs to prove one of the above cases so let us consider case (a). Consider a face f of T_C (or T_V). For sufficiency, suppose that only extreme-efficient DMUs in Z_C (or Z_V) are needed to describe f but that it is not an efficient facet. Then, if $(x_o, y_o) \in f \Rightarrow (x_o', y_o') \in f$,where $x_o' \geq x_o$ and $y_o' \leq y_o$ (i.e. (x_o', y_o') is a weakly-efficient bundle). In this case we have $(x_o', y_o') = (x_o, y_o) + \mu R$, $\mu > 0$ which is a contradiction. To establish necessity, suppose that both

[3] We will say that a set of extreme points and rays expresses/describes a facet when any production bundle on that facet can be described as a combination of these points and rays.

extreme-efficient DMUs and free-disposability rays are needed to describe f. Clearly in such a case f is a weakly-efficient facet. ■

Below we shall see that characterisations of the efficiency classification of faces as well as their dimensionality can be looked at from a dual point of view, i.e. by considering the weighting schemes associated with their supporting hyperplanes. In terms of efficiency classifications, and in parallel to the extensive discussion in the DEA literature on the efficiency classification of DMUs, we can claim the following:

Lemma 4 *Let f_a be a face f of T_C and f_b a face of T_V. Then*

(a) f_a is an efficient face if and only if $\exists\ H_a = \{(x,y) \in \mathbb{R}^{s+m} |\ -v_o x + u_o y = 0,\ (v_o, u_o) \in P_C\ ,\ v_o, u_o > \mathbf{0}\}$, *such that* $f = H_a \cap T_C\ (T_V)$.

(b) f_b is an efficient face if and only if $\exists\ H_b = \{(x,y) \in \mathbb{R}^{s+m} |\ -v_o x + u_o y = \beta_o,\ (v_o, u_o, \beta_o) \in P_V,\ v_o, u_o > \mathbf{0}\}$, *such that* $f = H_b \cap T_V$.

Proof. Consider the VRS case (the CRS case can be shown to be equivalent by setting $\beta = 0$). To establish necessity let us suppose that there exists an efficient face f which is only supported hyperplanes with at least one multiplier equal to zero. Without loss of generality we shall assume that $f = H^o \cap T_V$ where H^o is defined by a weighting scheme (v, u, β) such that $u > 0$ and $v = (0, v_2, ..., v_s)$, i.e. only the weight for input 1 is equal to zero. Let $(x_o, y_o) \in f$, and define $x_o^{'} = (x_{1o} + k, x_{2o}, ..., x_{so})$, $k > 0$. Clearly then $-vx_o^{'} + uy_o = \beta \Rightarrow (x_o^{'}, y_o) \in H^o$. In addition, from (2.13) and (2.18) we infer that $(x_o^{'}, y_o) \in T_V$. It follows then that $(x_o^{'}, y_o) \in H^o \cap T_V \Rightarrow (x_o^{'}, y_o) \in f$ which is a contradiction because $(x_o^{'}, y_o)$ is a clearly a weakly-efficient bundle.

For sufficiency suppose that $\exists\ H' = \{(x,y) \in \mathbb{R}^{s+m} |\ -vx + uy = \beta,\ (v, u, \beta) \in P_V,\ v, u > \mathbf{0}\}$ such that $f = H' \cap T_V$, but that f is not an efficient facet, i.e. if $(x_o, y_o) \in f \Rightarrow \exists (x_o^*, y_o) \in T_V$ where $x_o^* \leq x_o$. Clearly then, $-vx_o^* + uy_o > \beta$ which is a contradiction since H' is a supporting hyperplane of T_V. ■

Corollary 3 *A face f of T_C or T_V is a weakly-efficient face if and only if any hyperplane supporting it is defined by a weighting scheme which contains at least one zero multiplier.*

Note that all weakly-efficient facets of T_C (T_V) contain some Pareto-efficient production bundles. Formally:

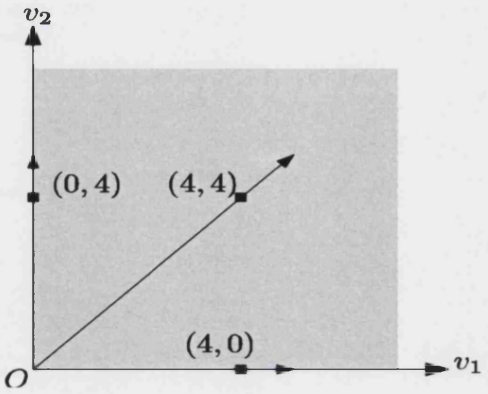


Figure 3.3: The intersection of P_{CD} with cuts $u = 3$, $\beta = 60$

Remark 2 *Let f_a be a facet of T_C and f_b a facet of T_V, and let $Eff(T_C)$, $Eff(T_V)$ denote the Pareto-efficient parts of T_C and T_V respectively. Then, $f_a \cap (Eff(T_C) \neq \varnothing$ and $f_b \cap (Eff(T_V) \neq \varnothing$.*

Proof. This is trivial since every facet will contain at least one extreme-efficient DMU. ■

Corollary 4 *Any weakly-efficient facet f of T_C (T_V) contains an efficient face of lower-dimension.*

For example, in figure 3.1 facets ABC and EGI are efficient but facet A$'$ACC$'$ is not as it is supported by hyperplane H_6 which has a zero multiplier for input $X1$. In addition, CD is clearly an efficient face contained in two weakly-efficient facets of T_V, although this is not immediately obvious in terms of the weighting schemes associated with its supporting hyperplanes (H_7 and H_8), which contain zero multipliers. To see that there is a supporting hyperplane for CD with all multipliers non-zero, consider the cone of optimal weighting schemes for face CD, $P_{CD} = \{(v, u, \beta) \in P_V | - vx_C + uy_C = \beta, -vx_D + uy_D = \beta\}$. The extreme rays of P_{CD}, are $C^1 = (4, 0, 3, 60)$ and $C^2 = (0, 4, 3, 60)$ and because in both these $u = 3$, $\beta = 60$, we can visualise this cone in two dimensions in figure 3.3. Now consider a weighting scheme $C^* = C^1 + C^2 = (4, 4, 6, 120)$. This is contained within P_{CD}[4] and hence also within P. It follows then that there exists a hyperplane with all multipliers non-zero ($H^* = \{(x, y) \in \mathbb{R}^3 | - 4X1 - 4X2 - 6Y = 120\}$) which supports face CD.

[4] Any non-negative combination of rays of a particular cone also lies within that cone. It follows then that in this case P_{CD} is the entire orthant.

3.3.2 Face-dimension

The preceding discussion leads naturally to the issue of the existence of well-defined rates of substitution raised by Olesen and Petersen[5]. Comparing the efficient face CD with the efficient facet ABC in figure 3.1, one can immediately observe that unlike for ABC there are two weighting schemes associated with face CD, namely H_7 and H_8. Each of these weighting schemes can be used to calculate *marginal rates of substitution/transformation* (MRS/MRT), along the boundary of the production set. The MRS corresponding to H_7 relate to rates of substitution along facet A$'$ACC$'$ and those corresponding to H_8 relate to movement within facet C$'$CDD$''$. However, zero multipliers are used to define both H_7 and H_8 and as a result their associated rates of substitution are not well-defined. Clearly face CD is contained in both these facets and hence the MRS within CD are not well-defined[6]. Let us note here that other than these two sets, there exist an infinite number of sets of MRS associated with the infinite number of hyperplanes supporting face CD, which can be expressed as nonnegative combinations of C^1 and C^2, like H^* above. However, one could argue against considering these, because such rates can relate to infeasible substitutions, i.e. substitutions which lead DMUs outside of the production possibility set. To guarantee feasible substitutions one needs to consider only the MRS calculated from weighting-scheme corresponding to facets of the polyhedral sets T_C and T_V, which are in fact extreme rays of their polar cones and correspond to *basic solutions* to the multiplier DEA LPs.

This is an important issue that has been overlooked by a lot of the DEA literature, mostly within the *weight-restrictions* related research, that assigns great importance to the selection of strictly positive weights for all inputs and outputs by all DMUs. Very recently, Cooper et al. (2006) addressed this problem by suggesting the selection of weighting schemes that have strictly positive multipliers but also maximum support from the dataset, and put forward an approach for this. We address this later.

Going back to the example in figure 3.1 there exists a unique strictly positive weighting scheme defining each efficient-facet (CD is not a facet as it is not fully-dimensional), and hence substitution

[5] Consider a facet f defined by weighting scheme (v, u, β) and define $F(x, y) = -vx + uy - \beta = 0$. The MRS between two inputs, say x_1 and x_2, is $-\frac{\partial F/\partial x2}{\partial F/\partial x1} = -\frac{v_2}{v_1}$. Similarly the MRS between y_1 and y_2, is $-\frac{\partial F/\partial y2}{\partial F/\partial y1} = -\frac{u_2}{u_1}$ and the MRT between x_1 and y_2, is $-\frac{\partial F/\partial y2}{\partial F/\partial x1} = \frac{u_2}{v_1}$.

[6] A set of rates of substitution/transformation along a segment of the frontier is well-defined if all rates of the set have non-zero and finite values. From the definition earlier it is obvious that not all MRS/MRT are well defined on weakly-efficient facets, since their defining weighting schemes contain zero multipliers.

within these is well-defined. On the other hand there exists an infinite number of hyperplanes with strictly positive multipliers containing face CD but on which infeasible substitutions exist and two supporting hyperplanes containing face CD on which substitutions are feasible but not well-defined.

Clearly then, there exists a link between the existence of well-defined rates of substitution and the dimension of the efficient faces of the production sets. We formalise tour observations below. First we consider the dimension of DEA production sets in general.

Lemma 5 *Let T_C, T_V be defined as in (2.13) and (2.18). Then,* $\dim(T_C) = \dim(T_V) = s + m$.

Proof. It follows directly from (2.13) and (2.18), or alternatively form the free-disposability postulate, that both T_C, and T_V contain interior points. The result follows. ■

Lemma 6 *Let $\bar{P}_C$, $\bar{P}_V$ be the sets of extreme rays of P_C, P_V and let I_C, I_V be their corresponding index sets. Then:*

(a) A face f of T_C is a fully-dimensional face if and only if there exists a unique $(v_o, u_o) \in \bar{P}_C$ such that $-v_o x + u_o y = 0 \ \forall (x, y) \in f$

(b) A face f of T_V is a fully-dimensional face if and only if there exists a unique $(v_o, u_o, \beta_o) \in \bar{P}_V$ such that $-v_o x + u_o y = \beta_o \ \forall (x, y) \in f$

Proof. Consider the VRS case (the CRS case is directly analogous). Clearly, any element of $\bar{P}_V$ defines a facet of T_V. Let $H_o = \{(x, y) \in \mathbb{R}^{s+m} | -v_o x + u_o y = \beta_o\}$. If f is uniquely supported by H_o then $\dim(f) = \dim(H_o \cap T_V) = s + m - 1$, i.e. f is a facet of T_V. On the other hand if f is supported by more than one hyperplanes defined by the extreme rays of P_V, say $H_1, H_2, ...H_k$ then $\dim(f) = \dim(H_1 \cap H_2 \cap ... \cap H_k \cap T_V) < s + m - 1$. ■

3.3.3 Combined results

We are now ready to combine previous results and obtain a classification scheme for fully-dimensional efficient faces. More specifically, by combining lemmata 4 and 6, we obtain the following:

Theorem 3 *A face f of T_C or T_V is a fully-dimensional efficient face, if and only if there exists:*

(a) a unique $(v_o, u_o) \in \bar{P}_C$, $v_o, u_o > 0$ such that $-v_o x + u_o y = 0 \ \forall (x, y) \in f$, in the CRS case.

(b) a unique $(v_o, u_o, \beta_o) \in \bar{P}_V$, $v_o, u_o > 0$ such that $-v_o x + u_o y = \beta_o \ \forall (x, y) \in f$, in the VRS case.

An immediate result is that we can identify fully-dimensional efficient faces by considering their defining weighting scheme and the number of extreme-efficient DMUs that they contain, i.e.:

Corollary 5 *A face f of T_C (T_V) is a fully-dimensional efficient face if and only if there exists a hyperplane H_o containing f, such that:*

(a) H_o is defined by a weighting scheme with all input-output multipliers positive.

(b) H_o contains at least $s+m-1$ $(s+m)$ extreme efficient DMUs.

In addition we can identify non fully-dimensional efficient faces as follows:

Corollary 6 *Let f_a be a non-fully-dimensional efficient face T_C and f_b a non-fully-dimensional efficient face T_V. Then:*

(a) f_a contains less than $s+m-1$ extreme-efficient DMUs.

(b) f_b contains less than $s+m$ extreme-efficient DMUs.

Turning back to efficient surfaces of the production possibility set, we can identify those that are not fully-dimensional as follows.

Corollary 7 *Let f be a non-fully-dimensional efficient surface of T_C (T_V). Then:*

(a) f is only contained in weakly-efficient faces of T_C (T_V).

(b) f contains fewer than $s+m-1$ $(s+m)$ extreme-efficient DMUs.

The above results establish, as was also claimed by Olesen and Petersen (1996), that the existence of well-defined rates of substitution along the strongly-efficient part of the DEA frontier, requires the existence of fully-dimensional efficient surfaces. In other words, well-defined rates of substitution exist when there is a correspondence between efficient surfaces and efficient facets of the DEA production possibility sets. Going back to our example, all efficient surfaces apart from CD have well defined MRS and are also facets of T_V. The efficient surface CD is a two-dimensional face of T_V for which no well-defined MRS exist which only correspond to feasible substitutions within T_V.

3.4 Counter-examples for two erroneous assertions

Having provided a theoretical framework for the classification of efficient faces and surfaces we are able to critically examine the existing literature on this problem. This leads to the discovery of two important inconsistencies which we address separately in this section.

3.4.1 Extreme-efficient DMUs

The first point we would like to address relates to the number of extreme-efficient DMUs contained in faces of the production possibility set. Frequently in the related literature, fully-dimensional efficient faces are identified as those containing $s+m-1$ extreme-efficient DMUs in the CRS case and $s+m$ in the VRS case. As we saw in the previous chapter, Olesen and Petersen (1996, 2003) and Raty (2002) develop optimisation procedures for the identification of all fully-dimensional efficient faces of the production possibility set in which the identified faces are constrained to support an appropriate minimum number of extreme-efficient DMUs. However, assessing the efficiency classification as well as the dimension of faces based on the number of extreme efficient DMUs that they contain *can be misleading*, as illustrated by the following counter-example.

Consider use the 2-input 1-output example dataset given in table 3.3. This generates the production possibility set plotted in figure 3.4, the facial description of which is given in table 3.4. Let us note first that all DMUs are extreme-efficient. Turning to the faces of the production set, we can observe easily that all fully-dimensional efficient faces, such as facets ABF and ACEF contain at least $s+m$ extreme-efficient DMUs. However, no such general statement can be made for weakly-efficient faces. Face A$'$ACC$'$ for example contains two extreme efficient DMUs. Face BB$'$JKDD$'$ on the other hand, described by H_8, contains four extreme efficient DMUs and any subset of $s+m$ of them can be seen to be affinely independent. The same is true for face H$'$HEFGII$'$.

In general, we know from the earlier results that any fully-dimensional efficient face of T_C (T_V) will contain at least $s+m-1$ ($s+m$) extreme-efficient units. However, we *cannot* claim that *any* face that contains that many extreme-efficient units is a fully-dimensional efficient face, as is done in much of the aforementioned literature. In other words, the existence of at least $s+m-1$ ($s+m$) extreme efficient units contained in a face of T_C (T_V) is a *necessary but not* also a *sufficient* condition for identifying that particular face as being fully-dimensional and efficient.

DMU	Input 1 ($X1$)	Input 2 ($X2$)	Output (Y)
A	2	4	3
B	4	2	3
C	2	8	5
D	8	2	5
E	4	6	7
F	4.5	4.5	7
G	6	4	7
H	3.5	9	7
I	9	3.5	7
J	5	2	4
K	6	2	4.5

Table 3.3: Example dataset

Supporting Hyperplane	Associated Weighting Scheme (v_1, v_2, u, β)	Spanning DMUs
H_1	$(5, 7, 5, -19)$	(B, F, J)
H_2	$(4, 4, 3 - 15)$	(A, B, F)
H_3	$(5, 13, 10, -11)$	(F, K, J)
H_4	$(1, 1, 0, -6)$	(A, B)
H_5	$(5, 15, 12, -6)$	(F, G, K)
H_6	$(3, 1, 2, -4)$	(A, C, E, F)
H_7	$(1, 0, 0, -2)$	(A, C)
H_8	$(0, 1, 0, -2)$	(B, D, K, J)
H_9	$(1, 5, 4, 2)$	(C, D, K)
H_{10}	$(6, 1, 5, 5)$	(C, E, H)
H_{11}	$(1, 6, 5, 5)$	(D, G, I)
H_{12}	$(0, 0, 1, 7)$	(E, F, G, H, I)
H_{13}	$(4, 0, 3, 7)$	(C, H)
	$(0, 4, 6, 7)$	(D, I)

Table 3.4: Hyperplanes and weighting schemes for the DMUs in table 3.3

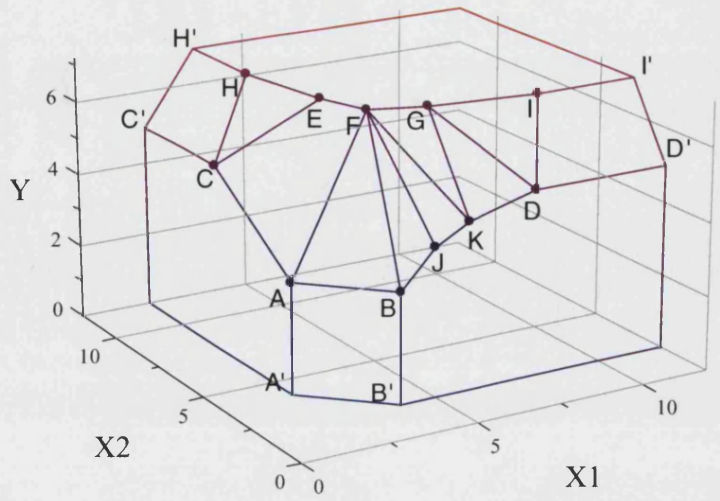


Figure 3.4: VRS production possibility set generated by the DMUs in table 3.3

3.4.2 Support of weighting-schemes

Finally, we address another point relating to the *support* of weighting schemes from the dataset, a concept used by Cooper et al. (2006). Applying the results from the previous section we shall show that their application of this concept is flawed.

The 'contact' between a weighting scheme's defining hyperplane and the production possibility set can be measured by considering the number of extreme-efficient DMUs contained in the supporting hyperplane defined by the weighting scheme $(v_t, u_t, \beta_t) \in P_V$, i.e. the number of DMUs $j \in \hat{E}_V$ that satisfy the equality $-v_t\mathbf{x}_j + u_t\mathbf{y}_j = \beta_t$. Formally, we consider the VRS case and define $\hat{E}_V(v_t, u_t, \beta_t)$ and $Supp(v_t, u_t, \beta_t)$as follows:

$$\hat{E}_V(v_t, u_t, \beta_t) = \{j \in \hat{E}_V | -v_t\mathbf{x}_j + u_t\mathbf{y}_j = \beta_t\} \tag{3.1}$$

$$Supp(v_t, u_t, \beta_t) = \left|\hat{E}_V(v_t, u_t, \beta_t)\right| \tag{3.2}$$

We refer to $Supp(v_t, u_t, \beta_t)$ as the support of weighting scheme (v_t, u_t, β_t) from the dataset. Consider the set of optimal weighting for a particular DMU $o \in J$. Following Cooper et al. (2006), we say that a weighting scheme in this set has maximum possible support from the dataset, if the

number of extreme-efficient DMUs contained in its defining hyperplane is greater than or equal to the numbers contained in the defining hyperplanes of all other weighting schemes in this set. Formally, we define $MaxSupp_j$ as follows:

$$MaxSupp_j = Max_i\{\hat{E}_V(v_i,u_i,\beta_i)|\ -v_i\mathbf{x}_j + u_i\mathbf{y}_j = \beta_i,\ (v_i,u_i,\beta_i) \in P_V\} \tag{3.3}$$

so that weighting scheme $(v_t,u_t,\beta_t) \in P_V$ has maximum possible support from the dataset if $\hat{E}_V(v_t,u_t,\beta_t) = MaxSupp_j$. Finally we define for every Pareto-efficient DMU j the subset $MaxP_V^j$ of its optimal weighting schemes, containing those that have maximum possible support form the dataset, i.e.:

$$MaxP_V^j = \{(v_i,u_i,\beta_i) \in P_V|\ -v_i\mathbf{x}_j + u_i\mathbf{y}_j = \beta_i, \hat{E}_V(v_t,u_t,\beta_t) = MaxSupp_j\}$$

Cooper et al. (2006) claim that: $\forall j \in Eff_V(J)$, $\exists(v_t,u_t,\beta_t) \in MaxP_V^j : (v_t,u_t) > \mathbf{0}$. In other words, among those optimal weighting schemes for Pareto-efficient DMU j that have maximum possible support from the dataset, there exists at least one weighting scheme that contains only positive multipliers. In addition, they introduce an optimisation procedure for the selection of such weighting schemes by all Pareto-efficient DMUs.

This claim is *incorrect.* We provide a counter-example for this in figure (3.4). For the Pareto-efficient DMU K the weighting scheme with maximum support corresponds to hyperplane H_8 that contains four extreme-efficient DMUs (B, D, K, J). Clearly this defines a weakly-efficient facet of T_V and its corresponding weighting scheme contains two zero multipliers[7]. The same observation holds for any extreme-efficient DMU contained in the weakly-efficient facets defined by hyperplanes H_8 and H_{11}.

[7]Note that we only need to consider facet-supporting hyperplanes of T_V. By definition hyperplanes that do not support facets will not have maximum possible support from the dataset.

Chapter 4

Sensitivity and Stability Analysis: existing approaches

4.1 Introduction

Within the DEA literature, the term *sensitivity analysis* refers to the family of approaches that study the sensitivity of the results obtained by DEA models to changes in model instances, such as alteration of the model specification, addition/exclusion of DMUs and data variations (or perturbations). A review of the relevant DEA literature reveals a keen interest in sensitivity analysis by many researchers in the field.

Of particular interest within this strand of DEA research is the sensitivity analysis related to perturbations in the data. In a practical context this kind of sensitivity analysis can be used to assess the efficiency of DMUs when their data change posterior to the initial DEA analysis, or, perhaps more importantly, to obtain an understanding of the impact of uncertainty with respect to data values. The focus of existing approaches is on the effect of data perturbations in the stability of DEA classifications, i.e. the classification of DMUs into efficient and inefficient. The more specific term *stability analysis,* is used to refer to this.

In this chapter we review existing DEA sensitivity analysis methods with respect to data perturbations, explain the cases that they study and address their strengths and weaknesses. Following

Cooper et al. (2001, 2004), we can distinguish between two types of approaches to sensitivity analysis. Approaches of the first type assume that perturbations occur only in the data for a specific DMU and study how these affect its own classification, while approaches of the second type study the sensitivity of a DMU's efficiency classification when the data for all DMUs change simultaneously. We shall use the terms *local* and *global perturbation approaches* to distinguish between the two aforementioned types. These two types of approaches have different philosophies for sensitivity analysis. Based on the models and general philosophy of these approaches we distinguish between: *algorithmic approaches*, *metric approaches*, *envelopment approaches* and *multiplier approaches*. We discuss all four types of approaches separately in the following four sections.

4.2 Algorithmic approaches

In the first paper on sensitivity analysis in DEA, Charnes et al. (1985) note that the traditional linear programming sensitivity analysis methods are not suited for performing sensitivity analysis in DEA. To overcome this they turn to the development of algorithmic approaches for sensitivity analysis, which built on earlier research on sensitivity analysis in linear programming by Charnes and Cooper (1968), a line of work that was continued in a series of papers by Charnes and Neralic (see Neralic 1997 for a discussion) and recently by Neralic and Wendell (2004). Essentially, these are based on selecting an optimal basis to a DEA LP and establishing the range of input-output variations that do not affect the optimality of this basis. However, an optimal basis is not necessarily unique and so the choice of basis can affect the sensitivity analysis results. A review of the algorithmic approaches is beyond the scope of this chapter. Instead we focus on other approaches to sensitivity analysis that have appeared in the literature.

4.3 Metric approaches

Charnes et al. (1992, 1996) avoid the use of algorithmic approaches by employing metric concepts. More specifically, these approaches utilise the concept of *vector norms* in order to identify regions within which data variations do not alter a DMU's classification. The identified regions are dependent on the choice of norm. To illustrate this avenue to sensitivity analysis we discuss two models

that are based on the *Chebychev norm* (or l_∞-norm):

$$\begin{aligned} \max \quad & \delta \\ s.t. \quad & \mathbf{X}\boldsymbol{\lambda} + e\mathbf{s}^- + e\delta = \mathbf{x}_o \\ & \mathbf{Y}\boldsymbol{\lambda} - e\mathbf{s}^+ - e\delta = \mathbf{y}_o \\ & e\boldsymbol{\lambda} = 1 \\ & \delta \geq 0,\ \boldsymbol{\lambda}, \mathbf{s}^-, \mathbf{s}^+ \geq \mathbf{0} \end{aligned} \tag{4.1}$$

$$\begin{aligned} \min \quad & \delta \\ s.t. \quad & \mathbf{X}_{-o}\boldsymbol{\lambda} + e\mathbf{s}^- - e\delta = \mathbf{x}_o \\ & \mathbf{Y}_{-o}\boldsymbol{\lambda} - e\mathbf{s}^+ + e\delta = \mathbf{y}_o \\ & e\boldsymbol{\lambda} = 1 \\ & \delta \geq 0,\ \boldsymbol{\lambda}, \mathbf{s}^-, \mathbf{s}^+ \geq \mathbf{0} \end{aligned} \tag{4.2}$$

Note that both models relate to a VRS technology. Model (4.1) is intended for DMUs classified as inefficient by DEA. Let $(\delta^*, \boldsymbol{\lambda}^*, \mathbf{s}^{-*}, \mathbf{s}^{+*})$ denote the optimal solution to this model. When the input levels of the inefficient DMU o decrease to $\mathbf{x}_o - e\delta^*$ and simultaneously the output levels increase to $\mathbf{y}_o + e\delta^*$, then DMU o switches its classification to efficient. Generalising this, as long as $\delta \in [0, \delta^*)$, no reclassification will occur from a simultaneous change in input and output levels to $\mathbf{x}_o - e\delta$ and $\mathbf{y}_o + e\delta$ respectively. In that sense the optimal value δ^* defines a symmetric region within which data variations will not affect the classification of DMU o and is referred to as the *radius of stability* for DMU o. Note that this model does not account for Pareto-efficiency unless the objective function is replaced by $\delta + \varepsilon(es^- + es^+)$ in which case the model can be solved by the usual DEA two-phase procedure.

The rationale for model (4.1) is reversed in (4.2) which is intended for DMUs classified as radially-efficient by DEA. Recall that the matrices $\mathbf{X}_{-o}$ and $\mathbf{Y}_{-o}$ contain all columns $\mathbf{x}_j$ apart from $\mathbf{x}_o$ and hence this model can be related to the super-efficiency models by Andersen and Petersen (1993). The model seeks to determine the minimum value of δ such that the production

bundle $(\mathbf{x}_o + e\delta,\ \mathbf{y}_o - e\delta)$ is efficient with respect to the production technology generated by DMU o's peers. Consider an extreme-efficient DMU o and let $(\delta', \boldsymbol{\lambda}', \mathbf{s}^{-\prime}, \mathbf{s}^{+\prime})$ denote the optimal solution to (4.2). As long as $\delta \in [0, \delta')$ then DMU o is guaranteed to remain Pareto-efficient following the perturbation[1]. If DMU o is either a) Pareto-efficient but not extreme, or b) weakly-efficient, then it can be shown that at the solution to (4.2) $\delta' = 0$, i.e. even an infinitesimal perturbation in input and/or output levels would alter the efficiency classification of DMU o.

Let us note here that in the discussion above we have assumed that inputs and outputs change by equal amounts. Charnes et al. (1992) also study the case where different inputs and outputs change by different amounts by assigning weights to variable δ in the above models. If $\mathbf{d}^- \in \mathbb{R}^s_+$ and $\mathbf{d}^- \in \mathbb{R}^m_+$ contain the weights for the inputs and outputs respectively, then overall the changes in the input output bundle for DMU o can be depicted as $(\mathbf{x}_o - \mathbf{d}^-\delta,\ \mathbf{y}_o + \mathbf{d}^+\delta)$ for model (4.1) and $(\mathbf{x}_o + \mathbf{d}^-\delta,\ \mathbf{y}_o - \mathbf{d}^+\delta)$ for model (4.2).

To illustrate how these models work consider the examples plotted in figures 4.1 and 4.2. Figure 4.1 illustrates the case of inefficient DMUs. The squares surrounding units E, D and F are generated by the Chebychev norms with radius δ^*. For example, the horizontal and vertical arrows from D portray the vectors $\mathbf{x}_D - e\delta^*$ and $\mathbf{y}_D + e\delta^*$ which are of length δ^*. This length is determined by the points at which the square centered at D intersects the efficient frontier. As a result, point $(\mathbf{x}_D - e\delta^*,\ \mathbf{y}_D + e\delta^*)$ signifies a point at which DMU D changes its classification from inefficient to efficient. We mentioned earlier that model (4.2) cannot account for Pareto-efficiency. This can be illustrated by DMU F. The value of δ^* obtained from the optimal solution to F's model (4.2) instance can be used to determine point F$'$ where the square centered at F intersects the weakly-efficient frontier. In other words F signifies a point at which DMU F changes its classification from inefficient to weakly-efficient. Overall then, the finite positive value δ^* can be used to determine regions within which DMUs D, E and F can move without affecting their original efficiency classifications. In figure 4.1 these are the shaded regions contained entirely within the three squares[2].

In figure 4.2 we illustrate the case of radially-efficient DMUs. The squares surrounding units B and C are generated by the Chebychev norms with radius δ'. The horizontal and vertical arrows

[1] If $\delta = \delta'$, then A Pareto-efficient DMU o is not guaranteed to remain Pareto-efficient and might switch to weakly-efficient.

[2] Note that in general these regions are not closed.

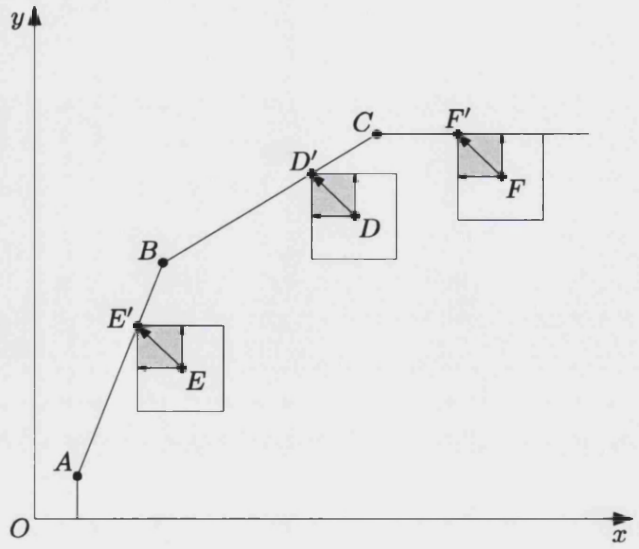


Figure 4.1: Metric stability regions for inefficient DMUs

from B portray the vectors $\mathbf{x}_B + e\delta'$ and $\mathbf{y}_B - e\delta'$ which are of length δ' and can be used to determine the point B′ at which the square centered at B intersects the efficient frontier of the production possibility set generated by B's peers. At this point DMU B changes its classification from efficient to inefficient. In general, given an optimal solution to (4.2), a stability region for an efficient DMU o can be determined by considering all possible production bundles ($\mathbf{x}_B + e\delta$, $\mathbf{y}_B - e\delta$) and allow δ to vary within the interval $[0, \delta')$. Like before, these are the shaded regions contained entirely within the three squares. For example, further changes in the input-output levels of DMU B, beyond those determined by point ($\mathbf{x}_B + e\delta'$, $\mathbf{y}_B - e\delta'$) will result in DMU B changing its classification from efficient to inefficient. There is an important reason why the interval for δ is open: if $\delta = \delta'$ and DMU o is a Pareto-efficient DMU, then the resulting production bundle ($\mathbf{x}_o + e\delta^{'}$, $\mathbf{y}_o - e\delta^{'}$) is not necessarily Pareto-efficient. This is illustrated by DMU C for which the projection point $C' = (\mathbf{x}_C + e\delta^{'}, \mathbf{y}_C - e\delta^{'})$ is clearly weakly-efficient. Finally notice that the radii of stability for DMUs D and E are zero. DMU D is a non extreme Pareto-efficient DMU while DMU E is a weakly-efficient DMU. This implies that these DMUs can be removed without affecting the shape of the production possibility set. Hence, be solving the instances of (4.2) for D and E one obtains that $\delta^{'} = 0$, i.e. even a minute change in the input-output levels of D and E can alter their efficiency classifications.

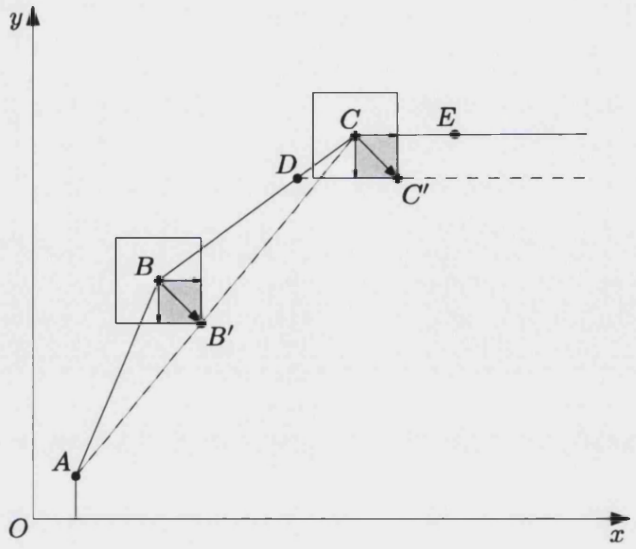


Figure 4.2: Metric stability regions for efficient DMUs

4.4 Envelopment approaches

Envelopment approaches are so named because they utilise DEA models in envelopment form to formulate the problem of sensitivity analysis of DEA efficiency classifications. Within this strand, both methods that deal with global and local perturbations have appeared. The local perturbation approaches work towards the identification of stability regions for individual DMUs. All of these utilise super-efficiency type models. Seiford and Zhu (1998b) consider separately the cases of increases in inputs and decreases in outputs, in a CCR model, and identify input and output stability regions respectively. However, the method described in Seiford and Zhu can cycle in certain cases, thus not allowing for the identification of stability regions. This is noted in Boljuncic (2006) who considers the case of simultaneous increases in inputs and decreases in outputs and identifies stability regions for DMUs within the BCC framework. We will discuss the theory behind these approaches and give an illustration of how they work with the aid of figures 4.3 and 4.4.

Consider the approach by Seiford an Zhu for identifying the input stability region of DMU $o \in J$. The method starts by solving an LP for every input. These LPs, developed by Zhu (1996),

are as follows:

$$\begin{aligned} \min \quad & \beta_k \ \forall k = (1, ..., s) \\ s.t. \quad & \textstyle\sum_j \lambda_j x_{kj} \leq \beta_k x_o \\ & \textstyle\sum_j \lambda_j x_{ij} \leq x_{io} \ \ \forall i \neq k \\ & \textstyle\sum_j \lambda_j y_{rj} \geq y_{ro} \ \ \forall r = 1, ..., m \\ & \lambda_j, \ \beta_k \geq 0 \end{aligned} \tag{4.3}$$

The optimal solutions to these s LPs can be used identify s separate projection points for DMU o on the efficient frontier of the production set generated by the rest of the DMUs. It is then checked whether the hyperplane generated by these s projection points (in input space only) defines a valid inequality for the reduced production possibility set i.e. if H is the supporting hyperplane, we check whether there exists a DMU $k \in (\ J \setminus o) : k \in H^{\geq}$. If so, then a stopping rule comes into effect and the input stability region scan be identified as the intersection of the halfspace $H^{\geq}$ with the cone from DMU o defined by the s free disposability rays indicating increases in inputs. If not, the following additional LP is solved:

$$\begin{aligned} \min \quad & \textstyle\sum_i \rho_i \\ s.t. \quad & \textstyle\sum_j \lambda_j x_{kj} \leq \rho_i x_{io} \ \ \forall i = 1, ..., s \\ & \textstyle\sum_j \lambda_j y_{rj} \geq y_{ro} \ \ \forall r = 1, ..., m \\ & \lambda_j, \ \rho_i \geq 0 \end{aligned} \tag{4.4}$$

This model projects DMU o on the reduced production set, by considering all inputs at the same time. Following this, by considering each component ρ_i separately, the projection point (which can be an existing DMU) is decomposed in s separate points indicating artificial DMUs generated by separate improvements in the levels of the s inputs of DMU o. The convex region defined by DMU o, its projection point and the s artificial DMUs is identified as a sub-region of the complete input stability region and we move to the next iteration where the entire aforementioned procedure is applied separately to each of the s artificial DMUs until the stopping criterion is met.

Hence, the complete input stability region is sequentially built by separate sub-regions identified in different iterations. At every iteration at least s LPs are solved (or m for output stability regions), each of which will at the worst case give rise to a further s LPs in the next iteration and this carries on, so it is easy to see that the number of LPs can quickly build up.

To illustrate this approach consider figure 4.3 where DMUs A-E utilise two inputs to produce one unit of a single output. We examine the input stability region for DMU B. At first two instances of LP model (4.3) are solved, one for each input, and two separate projection points are identified, i.e. B_1 and B_2. The line going through these points is not a supporting hyperplane to the reduced production possibility set so we now need to identify a new projection point for DMU B by considering all inputs at once and solving model (4.4). The projection point is DMU E and this is decomposed further into two artificial DMUs, i.e. b_1 and b_2. For both of these the aforementioned procedure is repeated, but before this we can also identify the square Bb_1Eb_2 as a sub-region of the complete stability region. For b_1 two instances of LP (4.3) are solved to identify the projection points B_1 and E, the line through which envelopes[3] all DMUs in $J \setminus B$ and hence the input stability region for b_1, the triangle b_1EB_1, can be identified as a sub-region of the complete input stability region for B. For point b_2 the two projection points E and B_2 do not define a valid inequality for the reduced production set so b_2 is projected on the frontier again (by solving model (4.4)) at point D and this gives rise to two new artificial DMUs, namely b_3 and b_4, and the sub-region $b_2b_3Db_4$. For both b_3 and b_4 their projection points define valid inequalities for the reduced production set so the two regions b_3B_2D and b_4DE are identified and the procedure stops since there are no more artificial DMUs to consider. Finally, the complete input stability region for DMU B can be identified as the union of the five sub-regions identified in the course of the above procedure.

The approach by Boljuncic (2006) is different, in that it identifies a single stability region for both inputs and outputs. To start with, the method identifies a unique projection point for DMU $o \in J$ on the efficient frontier of the reduced production possibility set by solving one LP. Boljuncic suggests that model (4.3) from Seiford and Zhu's approach can be used. From this we also obtain a set of weights that define the supporting hyperplane to the facet on which the point

[3] Given a hyperplane (in this case a line) $H = \{(x, y) \in \mathbb{R}^{s+m} | -vx + uy = \beta\}$, we will say that H *envelopes* all DMUs in a set A, if $-v\mathbf{x}_j + u\mathbf{y}_j \leq \beta \; \forall j \in A$.

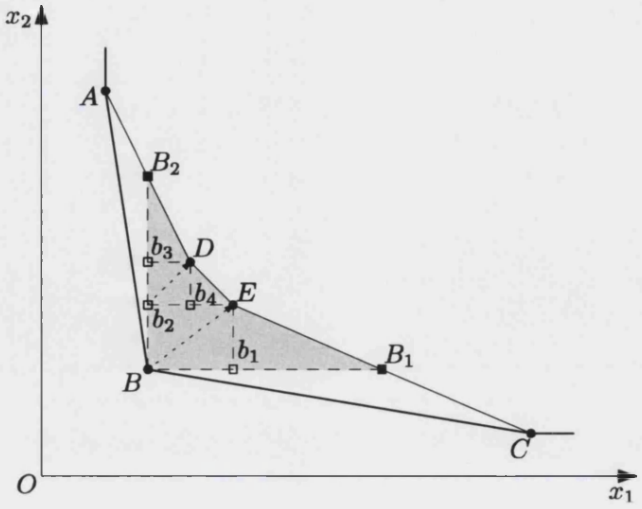


Figure 4.3: Input stability region for DMU B

is projected. Following this the method works towards identifying all relevant facets of the reduced production set required for the description of the stability region. The identification of new facets is done by applying parametric programming where the parameters correspond to input/output changes. Overall the approach calculates the acceptable changes in the inputs and outputs of the projection point such that the optimal basis of the original LP remains optimal. When changes exceed the acceptable range a pivot is performed and the new basis can be used to identify a new facet of the reduced production set. This is repeated until no further increases/decreases in inputs and/or outputs are possible at which point all relevant facets have been identified and the complete stability region can be identified.

To illustrate this approach consider figure 4.4. Suppose we wish to determine the stability region for DMU B. First, DMU B is projected on the efficient frontier of the reduced production possibility set, say at point B_1 (this could also be B_2), and this immediately identifies a facet of the reduced production possibility set, i.e. facet EC with supporting hyperplane H_3. We then need to determine the maximal input-output changes such that the projection point can be expressed as a combination of DMUs E and C, i.e. the initial basis remains optimal. We find that no further movement along facet EC is possible after point E. Thus we move to an adjacent facet of the reduced production possibility set by identifying an appropriate pivot, i.e. DMU C is replaced

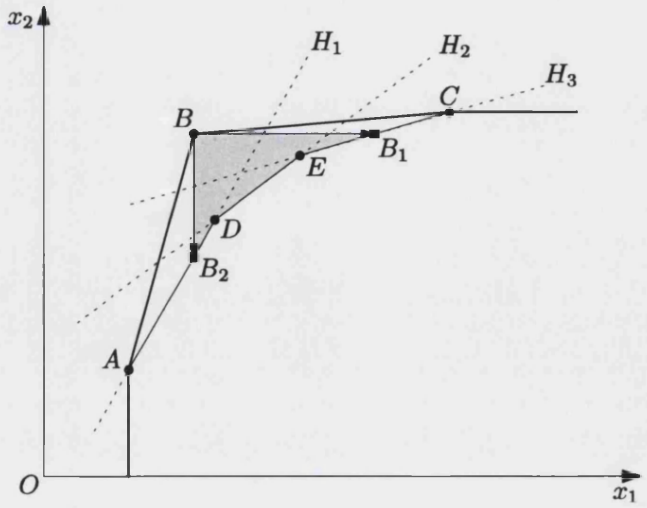


Figure 4.4: Input-Output stability region for DMU B

by DMU D and the new facet ED is identified. Repeating this we identify facet AD but we also find that no further increases in inputs and decreases in outputs are possible after point B_2 which lies on this facet. Hence there are no other feasible pivots and the procedure stops. The stability region can now be identified as the region BB_1EDB_2. Alternatively this can be described as the intersection of: a) the cone from B defined by the $s+m$ free disposability rays and b) the union of the three halfspaces $H_1^{\geq}$, $H_2^{\geq}$ and $H_3^{\geq}$.

Another set of envelopment approaches by Seiford and Zhu (1998a) and Zhu (2001), deal with the problem of global data perturbations where the data for all DMUs are allowed to vary simultaneously. To illustrate these we discuss the approach in Seiford and Zhu (1998a) who seek to determine the ranges in data variations when an efficient DMU o's inputs increase and the inputs for all other DMUs decrease, and simultaneously the converse is observed regarding the outputs. In general, data perturbations are modelled as follows:

$$\begin{aligned}
\hat{\mathbf{x}}_o &= \mathbf{x}_o + (\delta - 1)\mathbf{x}_o,\ \delta \geq 1 \\
\hat{\mathbf{y}}_o &= \mathbf{y}_o - (1 - \tau)\mathbf{y}_o,\ 0 < \tau \leq 1 \\
\hat{\mathbf{x}}_j &= \mathbf{x}_j - (\frac{\delta - 1}{\delta})\mathbf{x}_j\ \forall j \neq o \\
\hat{\mathbf{y}}_j &= \mathbf{x}_j + (\frac{1 - \tau}{\tau})\mathbf{y}_j\ \forall j \neq o
\end{aligned} \tag{4.5}$$

where $(\hat{\mathbf{x}}_j, \hat{\mathbf{y}}_j)\ \forall j \in J$, represent the resulting production bundles following the perturbations. More specifically, $(\delta - 1)$ and $(1 - \tau)$ represent the proportional increases/decreases in the input/output levels for DMU o, whereas $(\frac{\delta-1}{\delta})$ and $(\frac{1-\tau}{\tau})$ represent the proportional decreases/increases in the input/output levels for all other DMUs. The following model[4] is used to determine the ranges for parameters δ and τ under a CRS production technology:

$$\begin{aligned} \min \quad & \gamma \\ s.t. \quad & \mathbf{X}_{-o}\boldsymbol{\lambda} \leq (1+\gamma)\mathbf{x}_o \\ & \mathbf{Y}_{-o}\boldsymbol{\lambda} \geq (1-\gamma)\mathbf{y}_o \\ & \boldsymbol{\lambda} \geq \mathbf{0} \end{aligned} \tag{4.6}$$

Let $(\gamma^*, \boldsymbol{\lambda}^*)$ denote the optimal solution to (4.6). Seiford and Zhu prove that no changes occur in the efficiency classification of the originally efficient DMU o when the parameters δ and τ in (4.5) are allowed to vary in the ranges given in the following.

Theorem 4 *Consider an efficient DMU $o \in J$ and let the data perturbation for all DMUs be modelled by (4.5) using parameters δ and τ. If $1 \leq \delta \leq \sqrt{1+\gamma^*}$ and $\sqrt{1-\gamma^*} \leq \tau \leq 1$, where is the optimal value of the objective function in (4.6), then DMU o is guaranteed to remain efficient following the perturbations.*

Note that in the above all inputs and outputs are assumed to change by *equal proportions.* This is generalised by Zhu (2001) to the case where different factors are allowed to vary by different amounts. In addition, Seiford and Zhu (1998a) show that when (4.6) is infeasible, then DMU o remains efficient following any increase in its inputs and/or decrease in its output levels, modelled by (4.5).

4.5 Multiplier approaches

Another set of approaches by Thomson et al. (1994) and Thompson et al. (1996) also study proportional changes in the input-output levels of all DMUs and employ multiplier DEA models

[4] Let us also note that this is a simplified version of the model in Seiford and Zhu (1998a).

for determining the range in which these variations do not alter the efficiency classification of the particular DMU o being analysed. The focus in these methods is on extreme-efficient DMUs only. For any extreme-efficient DMU o there will always exist a valid set of input-output weights such that the hyperplane that they define contains DMU o only and envelopes all other DMUs. In other words, there exists a hyperplane which supports the production set at the vector associated with DMU o only. It follows that at this particular set of weights DMU o will achieve the highest efficiency score, in other words it will be *top-ranked.* Thomson et al. (1994) claim that in general, given a set of valid DEA weights, DMU o will remain top-ranked over some range of data variations and their approach (as well as the approach in Thompson et al. (1996)) works towards determining this range.

The procedure starts with the selection of a valid set of weights for an extreme-efficient DMU o and then the data for all DMUs are allowed to vary until DMU o is no longer top-ranked. For example to start with a 5% increase in the output levels of all other DMUs is considered and if DMU o is still top-ranked this is increased further until the situation is reversed at which point a maximum range is established. Thompson et al. (1996) report on experiments in which several different types of data variations are considered for all DMUs in the dataset.

Clearly, the type of data variations assumed can have an important impact on the established maximum range at which a DMU remains top-ranked with respect to a given set of weights. However, given a type of data variations, the selection of the weights on the basis of which DMUs are ranked can also affect the robustness of the established range for data variations. More specifically, given that extreme-efficient DMUs are top-ranked with respect to multiple sets of weights, selecting different weights will give rise to different (perhaps significantly) ranges of data variations within which extreme-efficient DMUs remain top-ranked. For this reason, Thompson et al. (1996) and Gonzalez-Lima et al. (1996) introduce the use of a special *interior-point algorithm* that selects weights close to what they call the *analytic center* of weights for DMU o and report that this increased the robustness of this approach when it was applied to particular datasets. Nevertheless, there is no guarantee that the selection of weights in this fashion establishes a range of variations close to the maximum possible range.

To illustrate this approach and its weaknesses consider figures 4.5a and 4.5b. The input-output data for the three DMUs, are: A= $(4,4)$, B= $(8,16)$ and C= $(20,16)$. We examine the changes in

the efficiency classification of DMU in more detail. For this purpose, we consider a 5% increase in output/increase in input for DMUs A and C, and the reverse for DMU B. These changes correspond to points A′ = (8.4, 15.2), B′ = (3.8, 4.2) and C′ = (19, 16.8). Following the procedure described earlier, we first need to establish a weighting scheme that defines a hyperplane that supports the production possibility set at point B only. Suppose that (v^*, u^*, β^*) is such a weighting scheme and that it defines hyperplane H^* in figure 4.5a. We can see clearly that following the 5% perturbations described earlier DMU B remains top ranked with respect to weighting scheme (v^*, u^*, β^*). Hence, we can now further increase the proportion at which perturbations occur, until DMU B is no longer top-ranked. At that point we have established the maximum range for the proportional perturbations. In figure 4.5b we illustrate how this approach is heavily dependent on the selected weighting scheme. Suppose that we had started by identifying weighting scheme (v', u', β'), which defines hyperplane H'. Following the 5% perturbation we observe that DMU B is no longer top-ranked, instead DMU C is now the top-ranked DMU. Hence, if we were to use weighting scheme (v', u', β'), we would establish a much smaller range for proportional data variations. The interior-point algorithm of Gonzalez-Lima et al. (1996) seeks to determine weighting schemes like (v^*, u^*, β^*), which give rise to greater ranges for data variations, but it offers no guarantee of establishing the maximum possible range.

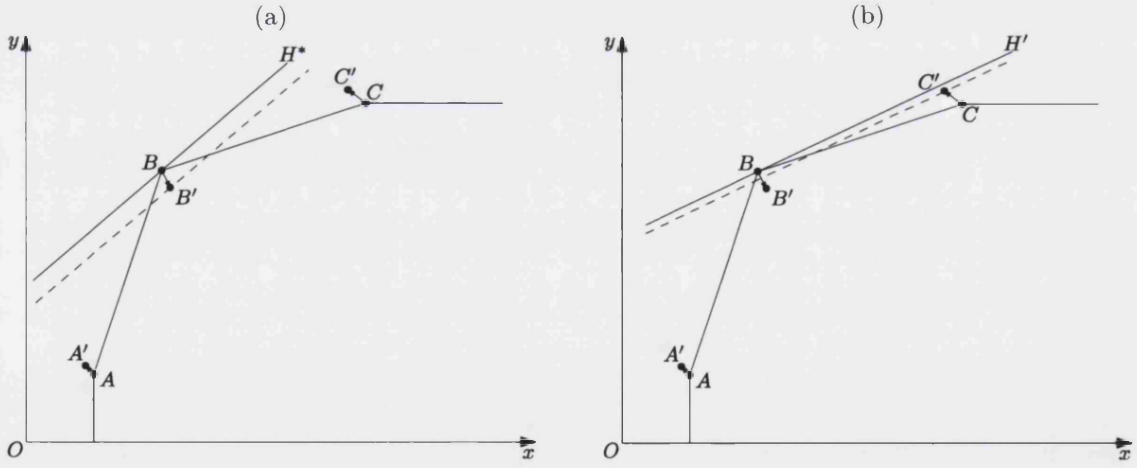


Figure 4.5: Multiplier approaches

4.6 An erroneous stability analysis approach

Finally, we report an approach by Jahanshahloo et al. (2005a) (a very similar approach also appears in Jahanshahloo et al. 2005b) who propose a characterisation of regions within which data perturbations do not affect any of the original efficiency classifications. We shall refer to these as *Conditional Stability Regions* (CSRs). Below we demonstrate that their theoretical results are erroneous.

To give a short description of their approach consider an efficient DMU $o \in J$. Essentially, the approach seeks to determine a region representing the geometrical 'difference' between the original production possibility set and a smaller set resulting from the exclusion of DMU o from J.

The process starts by identifying all facet-supporting hyperplanes passing through DMU o with respect to the original production possibility set. Let t be the number of identified hyperplanes. The i'th hyperplane supports a halfspace $H_i^- = \{(x,y) \in \mathbb{R}^{s+m} | -v_i x + u_i y \leq \beta_i\}$, where (v_i, u_i, β) is the defining weighting scheme. Following this we drop DMU o from the dataset. This results in a smaller production possibility set in which all originally efficient DMUs remain efficient but some originally inefficient DMUs could achieve efficient status. At this point we need to identify all originally inefficient DMUs that have now become efficient. This is done by repeating a DEA run for all originally inefficient DMUs. For each of these, exactly like for DMU o before, we identify all of their supporting hyperplanes with respect to the reduced production possibility set. Let r be the number of identified supporting hyperplanes. Now for each of these we define a set of halfspaces $H_i^+ = \{(x,y) \in \mathbb{R}^{s+m} | -v_i x + u_i y \geq \beta_i\}$.

At the end of this process we have identified two sets of halfspaces and we can now define two regions to respectively correspond to these sets, namely $S_1 = \bigcap_{i=1}^{t} H_i^-$ and $S_2 = \bigcup_{i=1}^{r} H_i^+$. Jahanshahloo et al. claim that the conditional stability region for DMU o can be defined as $S_1 \cap S_2$.

This claim is erroneous. Surprisingly, the error can be recognised easily even in the illustrations provided by Jahanshahloo et al.! We illustrate this with the example dataset in table 4.1 and figure 4.6. Suppose we are interested in determining the conditional stability region for DMU B. As described above, we first identify the hyperplanes supporting facets AB and BC. Following this we drop B from the dataset and identify the originally inefficient DMUs D and E that have now

DMU	A	B	C	D	E
Input	1	3	12	3	7
Output	1	8	9	5	8

Table 4.1: Counter-example data

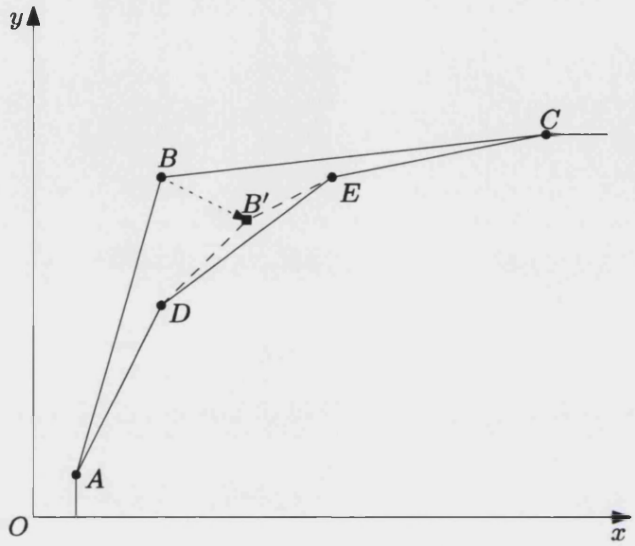


Figure 4.6: The erroneous region identified by Jahanshahloo et al.

switched to efficient. We then identify their supporting hyperplanes which respectively support facets AD, DE and EC of the new production set. Following the definition of the two separate regions above we can finally define B' CSR as the non-convex region ABCED.

This is obviously wrong. To establish this we perturb B to point B' with coordinates $(5, 7)$. Clearly this perturbation does not alter the classification of DMU B which remains efficient. However this is not the case for DMUs D and E which have been originally classified inefficient but have switched to efficient after the perturbation, since the frontier is now given by the segment ADB'EC.

In effect, Jahanshahloo et al. identify the region in which an efficient DMU can be perturbed without switching to inefficient. This region is obviously not 'conditional' as it does not guarantee the preservation of the efficiency classifications for other DMUs. Furthermore, their analysis is based on the implicit assumption that data perturbations will occur within the originally estimated production set and hence cannot account for arbitrary data perturbations.

Let us finally note that Jahanshahloo et al.'s approach is based on the explicit identification of the production possibility set which they propose to obtain by utilising the approach in Huang et al. (1997). As discussed in chapter 2 this approach cannot guarantee the identification of all supporting hyperplanes. Instead we would need to use one of the approaches that can perform this task described in that chapter .

Chapter 5

A new approach to stability analysis

5.1 Introduction

In this chapter we introduce a new approach to sensitivity/stability analysis in DEA. We are concerned with the characterisation of geometrical regions within which the data for one DMU can vary without affecting *any* original DMU classifications. The developments contained in this chapter utilise information obtained by the explicit identification of DEA production possibility sets, as described in chapter 2.

As we saw in the previous chapter, the use of such *stability regions* has already been reported in the literature. However, all related existing approaches only study particular types of data variations, e.g. increases in input levels and decreases in output levels. In addition to this, most existing methods focus on a single DMU at a time and study how variations in its data affect the stability of its own classification only.

In this chapter we take stability analysis two steps further; First, we do not constrain ourselves to certain types of data perturbations but study *arbitrary perturbations.* Second, because perturbations in a DMU's data do not necessarily affect only its own classification but possibly also those of its peers, we are concerned with identifying regions of stability within which *all* original DMU

efficiency classifications remain unaffected. We term this approach *Conditional Stability Analysis* (CSA).

The outline of the chapter is as follows. In section 5.2 we provide a preliminary discussion on arbitrary perturbations, illustrate the complexity of these with an example and point out that none of the existing approaches are powerful enough to study the effects of this type of data perturbations. Section 5.3 contains a set of new theoretical developments which we utilise in section 5.4 in order to provide exhaustive characerisations of *Conditional Stability Regions* for efficient and inefficient DMUs. We discuss how our new results on sensitivity analysis have opened interesting avenues for future research in the field in section 5.5, and in section 5.6 we address the computational difficulty of identifying Conditional Stability Regions by providing a computational framework that has the potential to reduce the overall effort considerably. We provide an illustrative example in section 5.7 and conclude in section 5.8.

5.2 Arbitrary Data Perturbations

We mentioned earlier that all existing approaches on sensitivity analysis study simplified versions of this problem. More specifically, they study different cases of data perturbations separately, mainly focusing on the case of increases in a DMU's input levels with simultaneous decreases in its outputs levels. Hence, there is a need to study other cases of data perturbations as well. What is also needed though, is the unification of all currently distinct cases under a general theoretical framework for sensitivity analysis. This of course is a much more complicated task; we shall see below that not all similar cases of data perturbations result in similar effects on efficiency classifications and that the results of current approaches are insufficient to model the complexity of arbitrary perturbations.

A general description of the problem is as follows: Consider a DMU $o \in J$ with data bundle $(x_o, y_o) \in \mathbb{R}^{s+m}$ and suppose that this bundle is now perturbed. Let $\hat{o}$ represent the perturbed DMU so that:

$$(\mathbf{x}_{\hat{o}}, \mathbf{y}_{\hat{o}}) = (\mathbf{x}_o, \mathbf{y}_o) + (\gamma, \delta),\ \gamma > -\mathbf{x}_o,\ \delta > -\mathbf{y}_o \tag{5.1}$$

Following this, what needs to be determined is whether this perturbation has altered the original

DMU	Input (x)	output (y)
A	2	2
B	4	6
C	8	9
D	12	10
E	6	7

Table 5.1: Example Dataset

classification of DMU o and, under CSA, any other DMU's classification. In general, we are interested in studying the conditions under which no alterations in efficiency classifications occur as a result of the perturbation in DMU o's data bundle.

To illustrate the many different types of effects that arbitrary perturbations have on the efficiency classifications of DMUs we use an example dataset given in table 5.1. We use the four panels of figure 5.1 to separately describe these effects. As can be observed from any of the four illustrations, of the five DMUs $A - D$ are efficient and E is inefficient. The efficient frontier is given by the piecewise linear segment $ABCD$.

We will consider data perturbations in DMU B's input-output data. For this simple two-dimensional case, we can consider four cases of perturbations, namely the four combinations of simultaneous increases or decreases in input and output levels. Consider the four orthants with B as the origin. Each of the four cases will lead to B moving within one of these orthants.

We start with the widely studied simplest case of perturbations, i.e. simultaneous increases in inputs and decreases in outputs. As illustrated in figure 5.1a this corresponds to moving B within the lower right orthant. Clearly, under this type of perturbations B will always remain within the original production possibility set. Hence, the perturbation is guaranteed not to increase B's efficiency score and not to reduce the efficiency score of any other DMU. For example if B moves to B_1 then it still remains efficient but now DMU E also becomes efficient and $AECD$ becomes the new efficient frontier. Further movement along the ray BB_1 will render DMU B inefficient, e.g. if it moves to B_2, $AECD$ becomes the efficient frontier. As we saw earlier Boljuncic (2006) studies an efficient DMUs 'region of efficiency' for this type of data perturbations within which it can move without becoming inefficient. In our example, B's region of efficiency is the region bounded by the lower right orthant and facet AE, the latter being part of the production possibility set resulting from the exclusion of B. We can modify this region to allow for conditional sensitivity analysis.

This will become obvious following the results in subsequent sections.

The other relatively simple case of perturbations regards simultaneous decreases in inputs and increases in outputs (upper left orthant of in 5.1b). Obviously, in this case a perturbed DMU which was originally efficient is guaranteed to remain efficient. On the other hand, if the perturbed DMU was originally inefficient, it might switch to efficient following the perturbation. Additionally, such perturbations may also reduce the efficiency scores of unperturbed DMUs and, as a result, force originally efficient DMUs to switch to inefficient. For example, moving the originally efficient B to point B_3, does not affect its own classification but results in DMU C becoming inefficient and the efficient frontier becoming AB_3D. Moving the originally inefficient DMU E to E_1 results in E switching to efficient. For inefficient DMUs facing this type of perturbations, we could use an approach in parallel to the approach by Boljuncic, and identify their region of inefficiency, both for the case of standard and conditional sensitivity analysis.

The remaining two cases concern simultaneous increases or decreases in both inputs and outputs, i.e. perturbing B within the upper right or lower left orthant. Under these types of perturbations the perturbed DMU is not guaranteed to always reside inside or outside the original production possibility set, making these cases less 'predictable' than the two previous ones. We illustrate the first of these two cases in figures 5.1c and 5.1d. In 5.1c, consider a simultaneous increase in both input and output levels for DMU B, so that it moves to point B_4 in the upper right orthant. This does not affect its own classification, however DMU C now becomes inefficient as the new frontier becomes AB_4D. Now contrast this with another possible movement of B within the same orthant, say to point B_5 in figure 5.1d. In this case DMU B becomes inefficient and the new frontier is given by ACD. Furthermore, we cannot conclude that moving B along ray BB_5 will always result in B becoming inefficient, as illustrated by point B_6. Moving B to B_6 does not alter its classification, as it remains efficient, but now DMU D switches to inefficient as the efficient frontier becomes $AECB_6$. Similar observations can be made for inefficient DMUs.

The above observations confirm that, even for a two-dimensional example, allowing for arbitrary data perturbations increases the complexity of possible effects on the classification of DMUs compared to the simplified cases that have already been explored. We must stress here that this complexity should increase further in higher dimensional cases, where some inputs and outputs might simultaneously increase and some others decrease, something that was not illustrated here.

Adding to this, studying the effects of data perturbations becomes even more complicated when the additional condition of preserving all DMUs' classifications is imposed. We undertake the task of studying such situations in the following sections.

5.3 Theoretical developments

In this section we develop the necessary theory for the characterisation of conditional stability regions. To that end, we consider a slightly different problem. More specifically, we work by looking at how the introduction of an additional DMU affects the original classifications of units. We shall see below how this allows us to arrive at more general results which we can utilise subsequently in characterising conditional stability regions. In what follows we consider the VRS case only[1], though it will be easy to see how all results can be adapted for the CRS case.

Let J' be the set of units obtained by the inclusion of an additional DMU k in set J, i.e. $J' = J \cup \{k\}$. We shall use dashed notation to define T', P', and E' that correspond to J', in exactly the same manner as was done for J in 2. We also define F_j (similarly F'_j) as follows:

$$F_j = \{(v, u, \beta) \in P |\ -v\mathbf{x}_j + u\mathbf{y}_j = \beta)\} \tag{5.2}$$

We refer to F_j as the efficiency cone for DMU j as it clearly contains all feasible weighting schemes that render it efficient. Note that F_j might be empty in which case j is an inefficient DMU. Of course, F_j, as defined above, relates to weak-efficiency and not Pareto-efficiency. We deliberately center our discussion around weak-efficiency to keep presentation as simple as possible. All results can be adapted to accommodate for Pareto-efficiency as well.

Recall that P' and F'_j (similarly P and F_j) are polyhedral cones and hence can be characterised by their extreme rays, as can be seen in the following general property (see e.g. Schrijver 2000):

Property 1 *Let C be a polyhedral cone and A the finite set of its extreme rays (v_i, u_i, β_i), $i \in B$*

[1] Since we only concern ourselves with the VRS cas we can simplify the notation by removing subscripts C and V, that were used to differentiate between the CRS and VRS cases. For example we now use T instead of T_V.

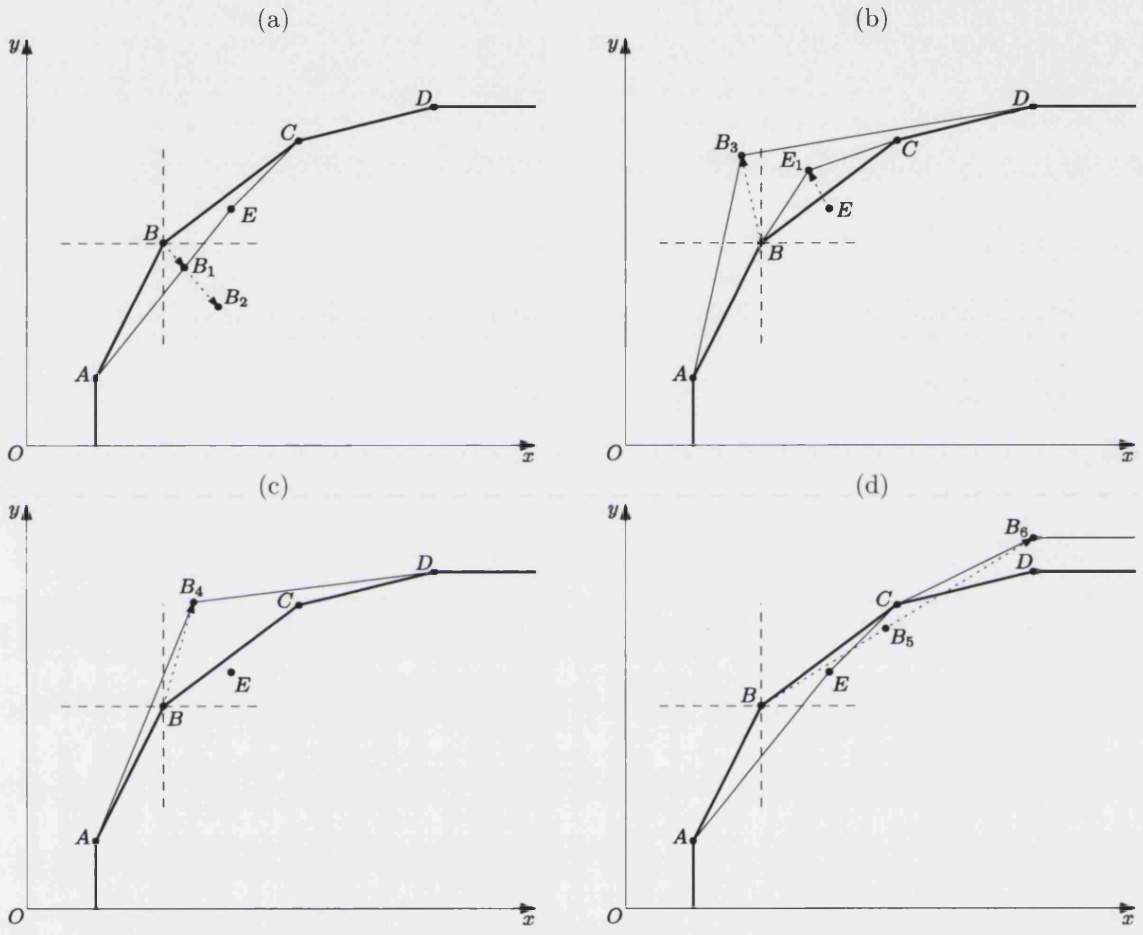


Figure 5.1: Perturbations and their possible effects

where B is an index set. Then:

$$\forall (v_o, u_o, \beta_o) \in C : (v_o, u_o, \beta_o) = \sum_{i \in B} \mu_i (v_i, u_i, \beta_i),\ \mu_i \geq 0, \sum_i \mu_i \neq 0$$

For each of the cones P, P', F_j, F_j', we denote their finite sets of extreme rays by $\bar{P}$, $\bar{P}'$, $\bar{F}_j$, $\bar{F}_j'$ respectively. We also use the index sets I, I', I_j, I_j' to correspond to the sets of extreme rays respectively.

The i'th extreme ray is associated with a hyperplane $H_i = \{(x, y) \in \mathbb{R}^{s+m} | - v_i x + u_i y = \beta\}$ which supports a facet of T (T'). The supporting hyperplane is in turn associated with a number of halfspaces. These are defined as follows:

Definition 19 *Consider the extreme rays (v_i, u_i, β_i), $i \in I$ of P (or any other of the above cones). We define $\forall i \in I$:*

$H_i^{\sim} = \{(x, y) \in \mathbb{R}^{s+m} | - v_i x + u_i y \sim \beta\}$, *where* $\sim = \{<, \leq, \geq, >\}$.

When the additional DMU k is introduced, this possibly affects the classification of some of the original DMUs in J. Below, we shall provide the necessary results for a characterisation of the resulting efficiency classifications of all original DMUs. In addition, we also examine the conditions that characterise k's classification.

First, consider the original DMUs in J. As established below, the introduction of the additional DMU k can only 'reduce' their efficiency cones.

Lemma 7 $\forall j \in J :\ F_j' \subseteq F_j$

Proof. $(v_o, u_o, \beta_o) \in F_j' \Rightarrow -v_o \mathbf{x}_j + u_o \mathbf{y}_j \leq \beta_o\ \forall j \in J'$ and since $J \subset J'$ it follows that $-v_o \mathbf{x}_j + u_o \mathbf{y}_j \leq \beta_o\ \forall j \in J \Rightarrow (v_o, u_o, \beta_o) \in F_j$. ■

A consequence of this is that all originally inefficient DMUs will remain inefficient, though some originally efficient DMUs might become inefficient. In addition, a necessary condition for a DMU $p \in J$ to be efficient with respect to the new production set, is that p was an originally efficient DMU. This is established in the following lemma:

Lemma 8 *Consider a DMU $p \in J$. The following hold:*

a) $p \notin E \Rightarrow p \notin E'$

b) $p \in E' \Rightarrow p \in E$

Proof.

a) $p \notin E \Rightarrow F_p = \varnothing \Rightarrow F_p' = \varnothing \Rightarrow p \notin E'$

b) $p \in E' \Rightarrow F_p' \neq \varnothing \Rightarrow F_p \neq \varnothing \Rightarrow p \in E$

■

Having established that only efficient DMUs may alter their classifications, what remains now is to characterise the conditions under which efficient DMUs remain efficient or switch to inefficient. The extreme rays of DMUs' efficiency cones are of crucial importance in obtaining this characterisation. In particular we shall utilise the following result.

Lemma 9 *Consider two DMUs* $p \in E$ *and* $k \notin J$. *Then,* $-v\mathbf{x}_k + u\mathbf{y}_k \sim \beta \; \forall (v,u,\beta) \in F_p$ *if and only if* $-v_i\mathbf{x}_k + u_i\mathbf{y}_k \sim \beta_i \forall (v_i,u_i,\beta_i) \in \bar{F}_p$.

Proof. Necessity is obvious since $\bar{F}_p \subseteq F_p$. Sufficiency follows directly from property (1). ■

The following lemma provides the necessary and sufficient conditions under which an originally efficient DMU in J preserves its efficiency classification after the introduction of the additional DMU k.

Lemma 10 *Consider two DMUs* $p, k: \; p \in E$ *and* $k \notin J$. *The following hold:*

a) $p \notin E'$ *if and only if* $-v_i\mathbf{x}_k + u_i\mathbf{y}_k > \beta_i \forall (v_i,u_i,\beta_i) \in \bar{F}_p$.

b) $p \in E'$ *if and only if* $\exists (v_o,u_o,\beta_o) \in \bar{F}_p : -v_o\mathbf{x}_k + u_o\mathbf{y}_k \leq \beta_o$.

Proof.

a) $p \notin E' \Rightarrow F_p' = \varnothing \Rightarrow -v\mathbf{x}_k + u\mathbf{y}_k > \beta \forall (v,u,\beta) \in F_p$. From lemma 3, this holds if an only if $-v_i\mathbf{x}_k + u_i\mathbf{y}_k > \beta_i \forall (v_i,u_i,\beta_i) \in \bar{F}_p$

b) Follows directly from part a.

■

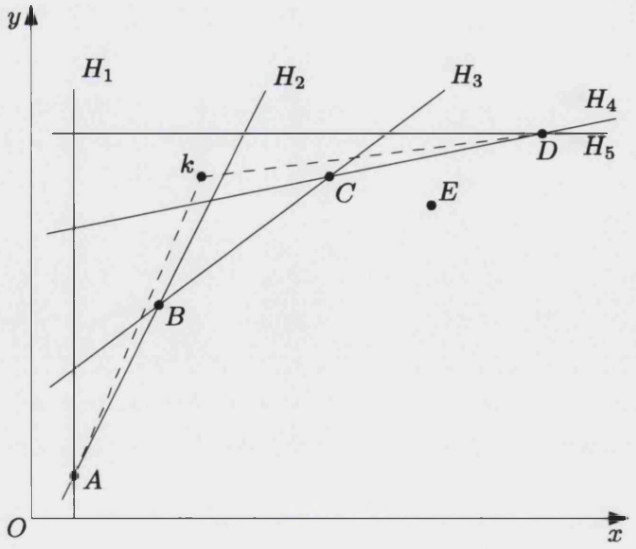


Figure 5.2: Effects of introducing an additional DMU

To illustrate the above, consider the example plotted in figure 5.2 where the original dataset consists of the five DMUs $A - E$. Now suppose that an additional DMU k is introduced to the dataset. According to the above, any weighting scheme contained in the new efficiency cones for any of the original DMUs must also be contained in their original efficiency cones. Clearly then, DMU E which is an originally inefficient DMU must remain inefficient as its original efficiency cone was empty to begin with. Now consider the remaining, originally efficient, DMUs. We can observe that the original production set is supported by five hyperplanes denoted $H_i,\ i = 1, ...5$. Following the introduction of DMU k, hyperplanes H_1 and H_5 envelope DMU k, i.e. $(x_k, y_k) \in H_1^{\leq}, H_5^{\leq}$. This implies that for DMUs A and D there exist weighting schemes in F_A and F_D able to envelope the additional DMU k and hence these DMUs will remain efficient with respect to the new production set T'. On the other hand, none of $H_2,\ H_3$ and H_4 envelope k as clearly $(x_k, y_k) \in H_2^{>}, H_3^{>}, H_4^{>}$. Because these are extreme rays of F_B and F_C, it follows from the above that there are no weighting schemes contained in these cones able to envelope k and hence these DMUs will become inefficient after the introduction of k.

We now turn our attention to the additional DMU k. Below we establish the conditions under which k will be classified as efficient/inefficient.

Lemma 11 *Consider a DMU k such that $k \in J'$, $k \notin J$. Then:*

a) $k \in E'$ if and only if $(x_k, y_k) \in \bigcup_{i \in I} H_i^{\geq}$.

b) $k \notin E'$ if and only if $(x_k, y_k) \in \bigcap_{i \in I} H_i^{<}$.

Proof. We need only prove part a. For necessity, suppose $(x_k, y_k) \notin \bigcup_{i \in I} H_i^{\geq}$. Then $\exists (v_o, u_o, \beta_o) \in P : -v_o \mathbf{x}_k + u_o \mathbf{y}_k < \beta_o \Rightarrow (x_k, y_k) \notin E \Rightarrow k \notin E'$ (from lemma 8).

For sufficiency: $(x_k, y_k) \in \bigcup_{i \in I} H_i^{\geq} \Rightarrow \exists (v_o, u_o, \beta_o) \in P : -v_o \mathbf{x}_k + u_o \mathbf{y}_k \geq \beta_o$ and $-v_o \mathbf{x}_j + u_o \mathbf{y}_j \leq \beta_o \forall j \in J$. Now let $\beta_o^* \geq \beta_o : -v_o \mathbf{x}_k + u_o \mathbf{y}_k = \beta_o^* \Rightarrow -v_o \mathbf{x}_j + u_o \mathbf{y}_j \leq \beta_o^* \forall j \in J' \Rightarrow (v_o, u_o, \beta_o) \in F_k' \Rightarrow k \in E'$. ■

For example, in figure 5.3 introducing a new DMU k at any point k_1 in the original production possibility set excluding the efficient frontier, will result in k being classified inefficient. In the case of weak-efficiency the boundary of the polyhedral production set serves as the efficient frontier, hence this region of inefficiency for k equates to the *relative interior*[2] of T (darkly shaded region). On the other hand introducing k at any point outside the original production set (k_3) as well as on its boundary (k_2) would render DMU k efficient (lightly shaded region). DMU k will be efficient if it is not enveloped by any of the original supporting hyperplanes. More specifically, k will be efficient if introduced in such a way that it is contained in any of the $H_i^{\geq}$ halfspaces corresponding to supporting hyperplane i, i.e. the union of all these halfspaces.

5.4 Characterisation of Conditional Stability Regions

5.4.1 General characterisation

The previous section established some general results on efficiency classification changes relating to the addition of a new DMU in the dataset. In this section we model data perturbations in the input-output levels of a particular DMU in such a way that the previous results can be brought into play.

Suppose we seek to identify the conditional stability region for DMU k. Recall that this is the region within which DMU k can be perturbed without affecting any of the original efficiency clas-

[2] The relative interior of T denoted $r.int(T)$ is defined as follws: $r.int(T) = \{(x, y) \in \mathbb{R}^{s+m} | -vx + uy < \beta \ \forall (v, u, \beta) \in P\}$.

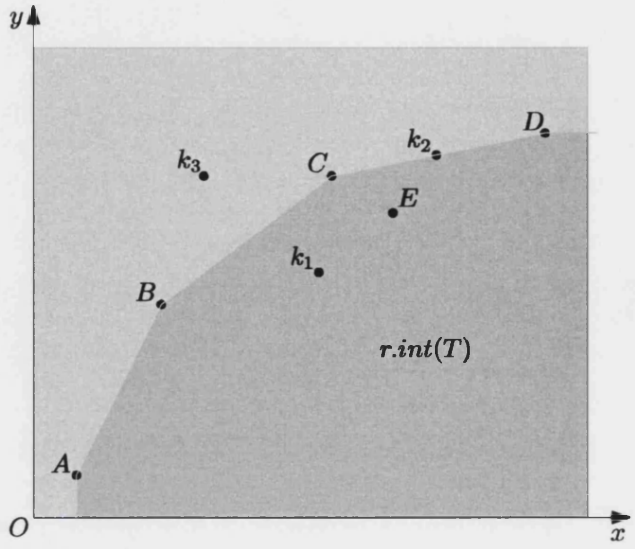


Figure 5.3: Classification of the additional DMU according to its position

sifications, including its own. In the spirit of many existing approaches we start by omitting DMU k from the dataset. Following this, a new set of efficiency classifications can be established. At this point, the identification of k's conditional stability region equates to the identification of the region within which the reintroduction of k will result in: a) k being classified identically to its original efficiency classification, and b) all other DMUs reverting to their own original classifications. In order to avoid the introduction of additional notation we assume henceforth that J' is the original set of DMUs and T' the original production possibility set.

First, consider the set E containing all DMUs declared efficient with respect to the production set generated by all DMUs except k. For every $j \in E$, we define a region K_j as follows:

$$K_j = \begin{cases} \bigcup\limits_{i \in I_j} H_i^{\leq}, \ j \in E \cap E' \\ \bigcap\limits_{i \in I_j} H_i^{>}, \ j \in E \setminus E' \end{cases} \tag{5.3}$$

K_j represents the region within which the reintroduction of DMU k will result in DMU j reverting to its original classification. For example, consider a DMU $j^* \in J'$ such that $j^* \in E \setminus E'$, i.e. j^* was originally inefficient but switched to efficient after the removal of DMU k from the

dataset. Since we want to 'force' j^* to revert to inefficient we need to reintroduce k in such a way that: $-v_o x_k + u_o y_k > \beta_o \forall (v_o, u_o, \beta_o) \in F_{j^*}$. According to lemma 10 this happens when $-v_i x_k + u_i y_k > \beta_i \forall (v_i, u_i, \beta_i) \in \bar{F}_{j^*}$, in other words only when $(x_k, y_k) \in \bigcap_{i \in I_{j^*}} H_i^{\geq} = K_{j^*}$.

Turning to DMU $k \in J' \setminus J$, we can define, in parallel to K_j, the region K_k as follows:

$$K_k = \begin{cases} \bigcup_{i \in I} H_i^{\geq}, \ k \in E' \\ \bigcap_{i \in I} H_i^{<}, \ k \notin E' \end{cases} \tag{5.4}$$

This definition guarantees that as long as the reintroduction of DMU k happens within the region K_k, k will maintain its original efficiency classification. For example, if $k \notin E'$, then we want to guarantee that it remains inefficient when reintroduced. To do so, we need to place k so that $-v_o x_k + u_o y_k < \beta_o \forall (v_o, u_o, \beta_o) \in P$. It follows that $(x_k, y_k) \in \bigcap_{i \in I} H_i^{<} = K_k$. In contrast if $k \in E'$, we need to make sure j remains efficient when reintroduced. For this to happen k would have to be placed on the boundary of, or exterior to the polyhedral production possibility set T, i.e. $(x_k, y_k) \in \bigcup_{i \in I} H_i^{\geq} = K_k$.

We now combine the results on regions K_j and K_k and obtain the following theorem, which characterises conditional stability regions for all DMUs.

Theorem 5 *Consider a DMU $k \in J'$. Its conditional stability region, denoted CSR_k, is given by:*

$$CSR_k = \bigcap_{j \in E} K_j \cap K_k \cap \mathbb{R}^{s+m} \tag{5.5}$$

Proof. First, note that CSR_k is not empty as it clearly contains at least one vector, namely the original production bundle (x_k, y_k). From lemmata 10 and 11 we know that DMU k as well as any other DMU in E will revert to their original classifications following a perturbation of k within CSR_k. The remaining set of DMUs $J \setminus E$ contains inefficient DMUs with respect to both production sets. It follows from lemma 8 that these are guaranteed to remain inefficient following any perturbation of DMU k. ■

5.4.2 Simplified cases

Having established a general characterisation of conditional stability regions, we now look at some particular cases where simpler characterisations can be obtained. These will be utilised in the following section where we consider the computation of conditional stability regions. First, consider the case where $k \in E'$ and also $E \setminus E' \neq \varnothing$. Then we have:

$$\bigcap_{j \in E \setminus E'} K_j = \bigcap_{j \in E \setminus E'} \bigcap_{i \in I_j} H_i^{>} \subseteq \bigcup_{i \in I} H_i^{\geq} = K_k \tag{5.6}$$

i.e. $\bigcap_{j \in E \setminus E'} K_j$ is a subset of K_k. This gives rise to a simpler characterisation of CSR_k as follows:

Corollary 8 *If $k \in E'$ and $E \setminus E' \neq \varnothing$, then $CSR_k = \bigcap_{j \in E} K_j \cap \mathbb{R}^{s+m}$*

In the following two cases we show that certain efficient DMUs and all inefficient DMUs respectively have identical stability regions for which simpler characterisations can be obtained. With respect to efficient DMUs, consider the following remark.

Remark 3 *Let $\hat{E}' \subseteq E'$ be the set of DMUs that serve as extreme points of the polyhedral production set T' and suppose DMU $k \notin \hat{E}'$. Then, $T' = T$.*

From this it follows that the conditional stability regions are identical for all efficient non-extreme DMUs. Formally:

Corollary 9 *For any pair of DMUs (k, l) : $k, l \in E' \setminus \hat{E}'$, the following holds:*

$$CSR_k = CSR_l \tag{5.7}$$

Now consider the case of inefficient DMUs. If $k \notin E'$ then clearly $E \setminus E' = \varnothing$, $\bigcap_{j \in E \setminus E'} K_j = \varnothing$ and hence $\bigcap_{j \in E} K_j = \bigcap_{j \in E \cap E'} K_j$. We can utilise this to obtain the following:

$$K_k = \bigcap_{i \in I} H_i^{<} \subseteq \bigcap_{i \in I} H_i^{\leq} \subseteq \bigcap_{j \in E} K_j \tag{5.8}$$

Since K_k is identical for all originally inefficient DMUs, it follows that all these have identical conditional stability regions, for which a simpler characterisation is given below:

Corollary 10 *For any pair of DMUs* (k,l) : $k,l \notin E'$, *the following holds:*

$$CSR_k = CSR_l = \bigcap_{i \in I} H_i^{\leq} \cap \mathbb{R}^{s+m} = r.int\,(T) \tag{5.9}$$

5.5 Future directions

The results presented so far have exhaustively characterised the problem of CSA. However, the research on this subject has opened some promising avenues for future research in the field of sensitivity analysis in DEA. We would like to devote this brief section to pointing out some possible directions for future research.

Firstly, we would like to note that whenever a DMU k is removed from set J', we have a chance to observe how this affects the classifications of other DMUs, i.e. set J. In the case that no additional DMUs become efficient, i.e. if $E = E'$, we can claim that the production set is stabile with respect to the removal of DMU k. In view of the earlier results, this will always be the case for inefficient as well as non extreme-efficient DMUs. However, this can also be the case for some extreme-efficient DMUs. Identifying the conditions under which this may happen could provide a link to sensitivity analysis with respect to addition/exclusion of DMUs. This is a fertile field for future research.

Another interesting problem that can be studied is the effect of perturbations on the magnitude of efficiency scores. At this stage, this is only possible for perturbed DMUs only. More specifically, it is possible to use the information on all available weighting schemes (of a full or a reduced production possibility set) to determine the efficiency score of the perturbed bundles. An open question for future research would be to generalise this to unperturbed DMUs, i.e. examine the effect of arbitrary perturbations for any type of DMU, on the efficiency scores of all DMUs in the dataset. The developments contained in this chapter have set the foundations for this kind of analysis.

5.6 Computational considerations

In this section we address some issues relating to the computation of conditional stability regions. In view of the earlier results we can base this on a procedure that involves the consecutive explicit identifications of production possibility sets. More specifically, this would start by identifying all supporting hyperplanes for the original production set T and following that the same would have to be done for every production set T' resulting from the removal of DMU $k \in E'$ from the dataset. Unfortunately, as we have seen in chapter 2 the explicit identification of one production set alone can be a computationally intensive process, hence the above would be no easy task. As a result, this procedure might only be applicable to relatively small datasets.

We would like to examine the computational burden of this approach in a relative context. The only comparable approaches to CSA are those by Seiford and Zhu (1998b) and Boljuncic (2006). Both these approaches compute stability regions and they do this by identifying relevant facets of the reduced production set (T in our case). In other words, they too work towards the explicit identification of production sets, only that they are interested in local parts of these. In addition, these approaches make no effort to connect the problem of identifying stability regions for different DMUs so that the same facets might have to be identified many times for different DMUs. Overall, this means that their computational burden should also be considerable although initially it might not seem so because of the use of LPs by these approaches.

For example consider the approach of Boljuncic (2006), as described earlier in figure 4.4. We can clearly see that for characterising the region of efficiency for DMU B, this approach would identify three facets of the reduced production possibility set, compared to a total of five facets that explicitly characterise the reduced production possibility set. Our approach would identify all of these five facets since it requires the explicit identification of the reduced production possibility set. Although in this example the difference in computational effort might seem minor, this is not generally the case. The additional effort required by our approach can be dramatically scaled up in larger examples.

5.6.1 Reducing computations

Fortunately, this situation can be improved by adapting the computational framework for CSA. More specifically, following the exclusion of one DMU from the dataset, the problem of explicitly characterising the resulting production possibility set T can be reduced to a smaller problem involving a subset of DMUs only. In effect, this is equivalent to focusing on local parts of the reduced production possibility set as is done by the approaches of Seiford and Zhu (1998b) and Boljuncic (2006). It is our belief that this modification has the potential to considerably reduce the overall computational effort.

In the absence of computational testing we cannot offer any guarantee that our approach is computationally comparable to the existing approaches. Nevertheless, we believe that any possible extra effort is worthwhile, given that CSA is more powerful than the aforementioned approaches, allowing, as it does, for arbitrary data perturbations and changes in the efficiency classifications of all DMUs.

Identifying the reduced production possibility set

We now describe the modified computational framework in more detail. A first thing to note is that the computational effort for explicitly characterising the original production possibility set T' can be reduced by only considering extreme-efficient DMUs (set $\hat{E}'$), as only these are necessary for generating T'. Of course this would require that we are able to distinguish between extreme and non-extreme-efficient DMUs. This can be done by the solution of one LP for each DMU. We could also use more efficient procedures described in Dula (1997) and Appa, Argyris, and Parthasarathy (2006), in which the LPs solved start small and grow larger with the identification of extreme-efficient units. Given the complexity of all existing algorithms for the generation of polyhedral sets, being able to reduce the number of data points (DMUs) can have a significant impact on the required computations. Following this, all DMUs can be scored with use of the obtained information on the supporting hyperplanes of T' (see also Briec and Leleu 2003; Appa and Williams 2006).

At this point we should be able to distinguish among extreme-efficient, efficient and inefficient DMUs. It is important to note that for the last two types no additional problem needs to be

solved. As established earlier, in both these cases respectively, all DMUs have identical conditional stability regions which can be readily obtained following the explicit characterisation of T'.

Now we address the identification of conditional stability regions for extreme-efficient DMUs. Again, let J' be the original set of DMUs for which we can define T', P', F'_j, etc. In addition, let E'_k be the set of units in J', that lie on at least one common facet with DMU k, i.e.[3]:

$$E'_k = \left\{ j \in J' | \exists\, (v_o, u_o, \beta_o) \in \bar{F}'_k : \; -v_o \mathbf{x}_j + u_o \mathbf{y}_j = \beta_o \right\} \tag{5.10}$$

To reduce the computational burden we first identify facets of the initial production set T' that remain facets of T. For this we shall utilise the following result, which establishes that the efficiency cones for all DMUs not in E'_k, are not affected by the removal of DMU k from the dataset.

Lemma 12 *For any DMU $j \in E' \setminus E'_k$, $F_j = F'_j$.*

Proof. From lemma 7 we know that $F'_j \subseteq F_j$ so we need only prove that $F_j \subseteq F'_j$. The polyhedral cone $F'_{j_o} \subseteq \mathbb{R}^{s+m+1}$ can be described as the intersection of a finite collection of halfspaces $-v\mathbf{x}_j + u\mathbf{y}_j \leq \beta$ which are defined by hyperplanes with normals $(-\mathbf{x}_j, \mathbf{y}_j, -1)$ where $(-\mathbf{x}_j, \mathbf{y}_j)$ correspond to DMUs in E'_{j_o}, i.e. the set of DMUs that co-span facets of T' with DMU j_o. Now by contradiction suppose that for a unit $j_o \in E' \setminus E'_k$, $F_{j_o} \not\subseteq F'_{j_o}$, so that there would exist $(v_o, u_o, \beta_o) \in F_k$ but $(v_o, u_o, \beta_o) \notin F'_k$. However, $(v_o, u_o, \beta_o) \in F_k \Rightarrow -v_o \mathbf{x}_j + u_o \mathbf{y}_j \leq \beta_o \; \forall j \in J$, but on the other hand if $(v_o, u_o, \beta_o) \notin F'_k$, i.e. (v_o, u_o, β_o) lies outside the cone F'_k, then for at least one of F'_k's supporting hyperplanes it will hold that $-v_o \mathbf{x}_j + u_o \mathbf{y}_j > \beta_o$, i.e. at least one of the DMUs in E'_{j_o} is not enveloped, which is a contradiction. Hence, $F_j \subseteq F'_j$. ■

From the above result we know that having explicitly identified T', we also hold explicit facial information for T, with respect to DMUs in set $\bar{E} = E' \setminus E'_k$. It follows that any 'new' facet of T can only be spanned by DMUs in $E \setminus \bar{E}$. Now consider the production set generated by the set of DMUs $E \setminus \bar{E}$, denoted $T(E \setminus \bar{E})$. We know that any 'remaining' facet of T must also be a facet of $T(E \setminus \bar{E})$. In other words $T(E \setminus \bar{E})$ is a subset of T which 'shares' some of its facets with T. This means that in order to identify the remaining facets of T we need only study the reduced problem of explicitly characterising $T(E \setminus \bar{E})$ and identify the relevant facets by inspection. Especially in

[3] Consider a face f of the production possibility set T (or T'). If k is an extreme-efficient DMU contained in f, we will say that k spans f.

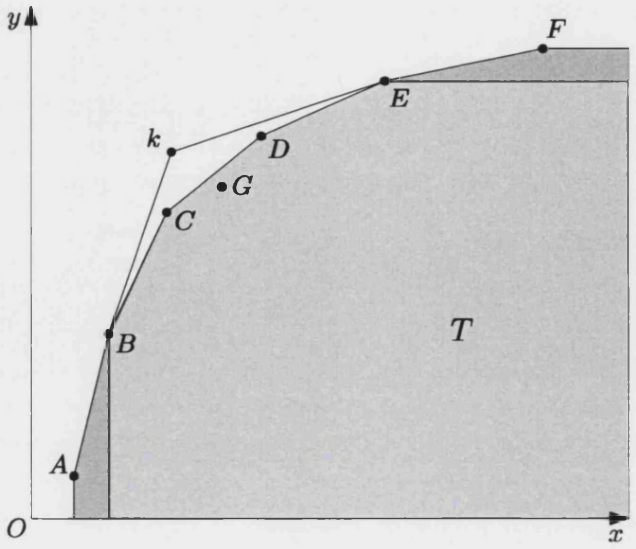


Figure 5.4: Identifying the facets of reduced problems

larger datasets where $E \setminus \bar{E}$ can be vastly smaller than E, this has the potential of considerably reducing the required computational effort.

Consider the example in figure 5.4. We are concerned with the explicit characterisation of production set T (darkly shaded region). Clearly $E' = \{A, B, k, E, F\}$ and $E = \{A, B, C, D, E, F\}$. Suppose DMU k is removed from the dataset. Since $E'_k = \{B, k, E\}$ and $\bar{E} = \{A, F\}$ we have: $E \setminus \bar{E} = \{B, C, D, E\}$. According to the above, we need only study the reduced problem of characterising the production set generated by $E \setminus \bar{E}$, denoted $T(E \setminus \bar{E})$ (the lightly shaded region entirely contained in T). Of course, in such a small example the computational savings are trivial. Nevertheless, we can clearly see from the illustration that the explicit characterisation of T requires seven facets of which four we can readily obtain from the information we hold about the original production set T' (facets AB and EF along with the weakly-efficient facets from A and F). The remaining three facets can be seen to also be facets of $T(E \setminus \bar{E})$ and can be obtained by explicitly characterising the latter. More specifically, following the explicit characterisation of $T(E \setminus \bar{E})$ the three relevant facets can easily be identified by inspection, i.e. simply by testing which of the facets of $T(E \setminus \bar{E})$ are able to envelope the remaining DMUs in E, i.e. those in set $\bar{E} = \{A, F\}$.

Identifying the new efficient DMUs

In the discussion above, the identification of set E necessarily precedes the explicit characterisation of T. Now we address this necessary first step of the 'reduced' procedure. Set E includes all originally efficient DMU bar k, i.e. the readily identifiable set $E' \setminus k$, along with the set $E \setminus E'$, i.e. the units that have switched to efficient following the removal of DMU k from the dataset. Clearly, $E \setminus E'$ can be identified by solving a DEA LP for all DMUs in $J \setminus E'$. However, this can be done more efficiently, by only considering a subset of $J \setminus E'$, as described below.

Lemma 13 *Consider a DMU $p \in J \setminus E'$. Then, $p \in E$ only if p is only dominated by facets of T' that are spanned by DMU k.*

Proof. Clearly any facet of T' spanned by DMUs in $E' \setminus E'_k$ will remain a facet of T and hence p will remain dominated. Hence, for p to switch to efficient it will have to be dominated only by facets spanning only by DMUs in E'_k. It follows directly that p must be dominated only by facets spanned by k. ■

Corollary 11 *Let θ_p^I and θ_p^O respectively be the input and output oriented efficiency scores of DMU p with respect to T'. Then $p \in E$ only if both $(\theta_p^I \mathbf{x}_p, \mathbf{y}_p)$ and $(\mathbf{x}_p, \theta_p^O \mathbf{y}_p)$ only lie on facets of T' spanned by DMU k.*

Note that the above only provide necessary but not sufficient conditions. To illustrate this consider DMU G in figure 5.4. Both of G's input and output oriented radial projections will lie on facets spanned by DMU k. However, DMU G clearly remains inefficient following the removal of DMU k from the dataset. Nevertheless, it follows from corollary 11 that in identifying which originally inefficient DMUs switch to efficient after the removal of DMU k, we need only examine DMUs for which DMU k can serve as a comparator DMU, i.e. spans the facet on which their targets lie. To that extent, we first evaluate efficiency scores for all DMUs with respect to all supporting hyperplanes[4] and for every DMU identify the maximum of these for both input and output orientations. We then need only solve a DEA LP for DMUs in a set S_k, defined to containi all DMUs for which both maxima are attained on facets spanned by DMU k. Again, this has the potential of significantly reducing the extra LPs that need to be solved.

[4] These are the cross-efficiencies as we will describe in the next chapter.

Finally, note that instead of identifying set E prior to the explicit characterisation of T we could also work towards the identification of $\hat{E}$, i.e. the extreme-efficient units of T, as was done for T' earlier. Given that these are the only necessary units for the generation of T, i.e. $T(\hat{E} \setminus \bar{E}) = T(E \setminus \bar{E})$, this would be preferable. We know that any DMU in $\hat{E}' \setminus k$ will remain extreme-efficient so we need to identify DMUs originally in $J \setminus \hat{E}'$ that become extreme-efficient following the removal of DMU k from the dataset. These new extreme-efficient DMUs are a subset of the new efficient DMUs which are all contained in set S_k as described earlier. Hence we know that all extreme-efficient DMUs in $\hat{E}$ are contained in $S_k \cup (\hat{E}' \setminus k)$ and we can identify them by applying the algorithms described in Dula (1997) or Appa, Argyris, and Parthasarathy (2006). In addition, we already know some of the extreme-efficient DMUs, namely those contained n $\hat{E}' \setminus k$, so we can initialise these procedures with this set. As was the case for T' earlier, this should allow for savings in computations for explicitly characterising T.

5.6.2 An algorithm for CSA

Here we give a detailed description of the suggested computational procedure for CSA. First, we would like to note the following:

(1) The procedure is based on the identification of extreme-efficient units prior to the explicit characterisation of various production sets

(2) $\bar{P}(\hat{E}')$ denotes the extreme rays of the polar cone $P(\hat{E}')$ corresponding to production set $T(\hat{E}')$. Since $\hat{E}'$ is the subset of extreme-efficient units of J, then clearly $T(\hat{E}') = T$ and so $\bar{P}(\hat{E}') = \bar{P}$. Similarly $\bar{P}(\hat{E}' \setminus \bar{E})$ corresponds to $T(\hat{E} \setminus \bar{E})$, and $\bar{I}_p$ is the index set for the elements $(v_i, u_i, \beta_i) \in \bar{P}(\hat{E}' \setminus \bar{E})$.

(3) S_k contains the set of DMUs in $J \setminus E'$ which are dominated only by facets of T' spanned by DMU k. This is calculated according to corollary 11. The DMUs in S_k that are also extreme-efficient with respect to $T(\hat{E})$ (or simply T), i.e. $S_k \cap \hat{E}$, can be identified by the solution of one LP per DMU, or by the procedures described in Dula (1997) or Appa, Argyris, and Parthasarathy (2006) as discussed earlier.

Algorithm 1 *Identify_CSRs* $\left(J',\ \hat{E}',\ CSR[|J|]\right)$

$Identify\ \bar{P}\left(\hat{E}'\right)$

$Identify\ E'$

$FOR\ \left(k \in J' : k \in E' \setminus \hat{E}'\right)$

$$CSR[k] = \left(\bigcap_{j \in \hat{E}} \bigcup_{i \in I_j} H_i^{\leq}\right) \cap \mathbb{R}^{s+m}$$

$FOR\ \left(k \in J' : k \in J' \setminus E'\right)$

$$CSR[k] = \left(\bigcap_{i \in I} H_i^{<}\right) \cap \mathbb{R}^{s+m}$$

$FOR\ \left(k \in J' : k \in \hat{E}'\right)$

$$K_k = \bigcup_{i \in I} H_i^{\geq}$$

$$J' = J \setminus \{k\}$$

$$\bar{E} = \hat{E}' \setminus E'_k$$

$$P_1 = \{(v, u, \beta) \in \bar{P}\left(\hat{E}'\right) \mid\ (v, u, \beta) \in \bar{F}'_j,\ j \in \hat{E}' \setminus E'_k\}$$

$$P_2 = \varnothing$$

$Identify\ S_k$

$Identify\ S_k \cap \hat{E}$

$$\hat{E} = \left(\hat{E}' \setminus \{k\}\right) \cup \left(S_k \cap \hat{E}\right)$$

$Identify\ \bar{P}\left(\hat{E} \setminus \bar{E}\right)$

$FOR\ (i \in \bar{I}_p)$

$IF\ -v_i \mathbf{x}_j + u_i \mathbf{y}_j \leq \beta_i\ \forall j \in \bar{E}$

$$P_2 = P_2 \cup \{(v_i, u_i, \beta_i)\}$$

$$\bar{P}\left(\hat{E}\right) = P_1 \cup P_2$$

$Identify\ E$

$FOR\ (j \in J :\ j \in E)$

$IF\ \left(j \in E \cap E'\right)$

$$K_j = \bigcup_{i \in I_j} H_i^{\leq}$$

$ELSE$

$$K_j = \bigcap_{i \in I_j} H_i^{>}$$

$$K_k = \bigcup_{i \in I} H_i^{\geq}$$

$$CSR[k] = \bigcap_{j \in E} K_j \cap K_k \cap \mathbb{R}^{s+m}$$

5.7 Illustrative example

To illustrate the ideas of the previous sections we use a small example where dataset J' consists of the fifteen units given in table 5.2. Production possibility sets T' and T are plotted in figure 5.5.

We start by identifying all extreme-efficient units, i.e. DMUs $B - E$, by using approaches described earlier. We then need to explicitly characterise T' which can be done by following any of the procedures described in section 2.4.2. In particular, we have used the approach of Appa and Williams (2006). With the obtained information on all supporting hyperplanes it is easy to establish subsequently that $E' = \{A, B, C, D, E, F\}$, as well as calculate the efficiency scores for inefficient units. Following this, we can immediately characterise CSRs for the inefficient DMUs $G - O$ and the non extreme-efficient DMUs A and F

To illustrate the computation of CSRs for extreme-efficient DMUs, let us consider the identification of the CSR for DMU $C \in J'$.We first drop C from the dataset (which leads to $E'_C = \{B, C, D\}$) and proceed explicitly characterise the production set T generated by the remaining units. Following the procedure described in the previous section, we know that any facet of T' spanned only by units in $\bar{E} = E' \setminus E'_C = \{A, E, F\}$ will also be a facet of T and that the remaining facets of T can be identified by solving the reduced problem of explicitly characterising $T\left(E \setminus \bar{E}\right)$. Before this is done it is necessary to identify set E. Clearly, any DMU in $E' \setminus k = \{A, B, D, E, F\}$ will belong to E so it remains to identify the 'new' efficient DMUs, i.e. set $E \setminus E'$. According to earlier results, this can be done by solving an LP only for DMUs with input and output-oriented efficiency scores on facets spanned by DMU k, i.e. DMUs in set $S_k = \{G, H, I\}$ (instead of all DMUs in $J \setminus E'$)[5]. This leads to $E \setminus E' = \{G, H\}$ and hence $E = \{A, B, G, H, D, E, F\}$. With all this information at hand we can proceed to explicitly characterise $T\left(E \setminus \bar{E}\right)$ and identify the relevant 'remaining' facets for the characterisation of T. Let us note again that instead of identifying all efficient units in T it would be preferable to only consider the extreme-efficient units, as was done at the beginning of the procedure for identifying T'.

At the end of this process we find that six hyperplanes are needed to characterise the new

[5] So we only solve three as opposed to nine LPs.

production possibility set T, more specifically $T = \bigcap_{i=1}^{6} H_i^{\leq}$. The hyperplanes along with their corresponding weighting schemes (w_i) are given in table 5.3. $\bar{F}_j$ and K_j, for all DMUs, are given in table 5.4. We now identify C's conditional stability region by considering its reintroduction to the dataset.

Notice that $E \setminus E' \neq \varnothing$, hence we know, from Corollary 1, that we can identify $C's$ conditional stability region as follows:

$$\begin{aligned} CSR_C &= \bigcap_{j \in E} K_j \cap \mathbb{R}^{s+m} = K_A \cap K_B \cap K_D \cap K_E \cap K_F \cap K_G \cap K_H \cap \mathbb{R}^2 \\ &= H_1^{\leq} \cap (H_1^{\leq} \cup H_2^{\leq}) \cap (H_4^{\leq} \cup H_5^{\leq}) \cap (H_5^{\leq} \cup H_6^{\leq}) \cap H_6^{\leq} \cap (H_2^{>} \cap H_3^{>}) \cap (H_3^{>} \cap H_4^{>}) \end{aligned}$$

As this is a simple example we can establish visually that $K_C \cap K_D = H_2^{>} \cap H_4^{>}$, and finally that:

$$CSR_C = H_1^{\leq} \cap H_2^{>} \cap H_4^{>} \cap H_5^{\leq}$$

On the graph, the conditional stability region for DMU C is the dark-shaded region $c_1c_2c_3c_4$. Note however that this does not include the line segments c_1c_2 and c_1c_4.

By repeating this process for all extreme-efficient DMUs we can calculate their CSRs. At the end of this procedure we have calculated all CSRs for all DMUs in the dataset. These are given in table 5.5.

5.8 Concluding remarks

In this chapter we presented an important extension to the study of sensitivity/stability analysis in DEA by introducing the concept of conditional stability. Contrary to prior approaches, CSA does not only study the stability of one DMU's efficiency classification to data perturbations, but the stability of the classifications of all DMUs. This gives CSA a strong practical appeal. We provided a thorough treatise of the theoretical framework for this problem and, by using results relating to the polyhedral nature of DEA production sets, we arrived at a complete characterisation of conditional stability regions, i.e. regions within which DMUs can be perturbed without affecting the efficiency classifications of their peers. Our results not only apply to standard CRS and VRS models, but

DMU	Input (x)	Output (y)
A	2	1
B	2	3
C	3	9
D	9	11
E	14	12
F	18	12
G	3	7
H	5	9
I	5	5
J	10	5
K	13	7
L	12	8
M	14	8
N	14	10
O	15	6

Table 5.2: Example dataset

i	H_i	$w_i = (v_i, u_i, \beta_i)$
1	$-x = 2$	$(1, 0, 2)$
2	$-4x + y = -5$	$(4, 1, 5)$
3	$-x + y = 4$	$(1, 1, 4)$
4	$-x + 2y = 13$	$(1, 2, 13)$
5	$-x + 5y = 46$	$(1, 5, 46)$
6	$y = 12$	$(0, 1, 12)$

Table 5.3: Hyperplanes and weighting schemes for the DMUs in table 5.2

DMU	F_j	K_j
A	w_1	$(x, y) \in \mathbb{R}^2 : \ x \geq 2$
B	w_1, w_2	$(x, y) \in \mathbb{R}^2 : (x \geq 2) \cup (-4x + y \leq -5)$
G	w_2, w_3	$(x, y) \in \mathbb{R}^2 : (-4x + y > -5) \cap (-x + y > 4)$
H	w_3, w_4	$(x, y) \in \mathbb{R}^2 : (-x + y > 4) \cap (-x + 2y > 13)$
D	w_4, w_5	$(x, y) \in \mathbb{R}^2 : (-x + 2y \leq 13) \cup (-x + 5y \leq 46)$
E	w_5, w_6	$(x, y) \in \mathbb{R}^2 : (-x + 5y \leq 46) \cup (y \leq 12)$
F	w_6	$(x, y) \in \mathbb{R}^2 : \ y \leq 12$

Table 5.4: Results

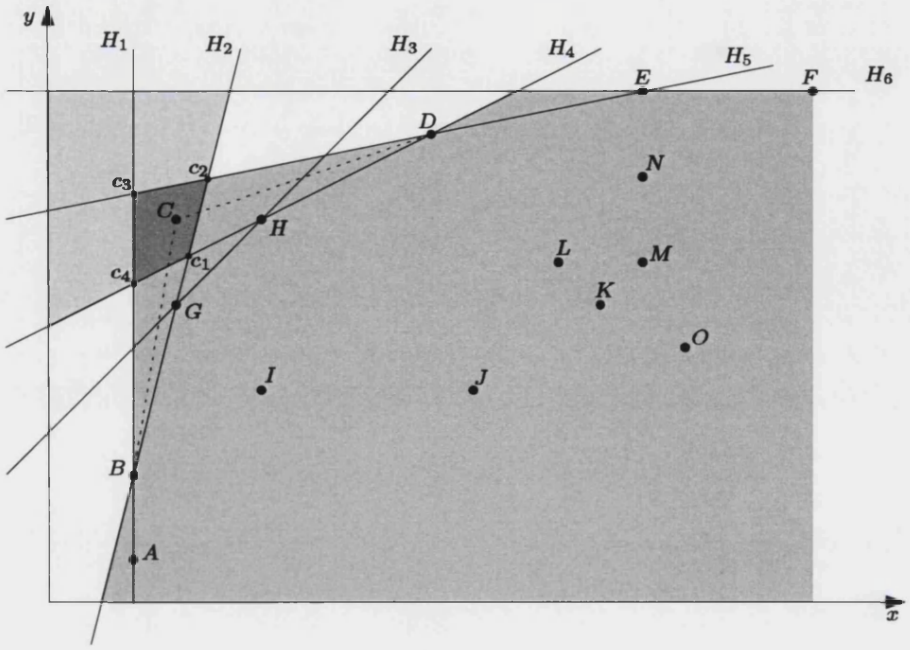


Figure 5.5: Conditional stability region

DMUs	CSR
$A,\ F$	$\{((x,y)\in\mathbb{R}^2_+ \mid (x\le 2)\cup(-6x+y\ge -9)\cup(-x+3y\ge 24)\cup(-x+5y\ge 46)\cup(y\ge 12))$ $\cap(x\ge 2)\cap((x\ge 2)\cup(-6x+y\le -9))\cap((-6x+y\le -9)\cup(-x+3y\le 24))$ $\cap((-x+3y\le 24)\cup(-x+5y\le 46))\cap((-x+5y\le 46)\cup(y\le 12))\cap(y\le 12)\}$
B	$\{(x,y)\in\mathbb{R}^2_+ \mid (x\le 2)\cup(-8x+y\ge 15)\cup(-x+3y\ge 24)\cup(-x+5y\ge 46)\cup(y\ge 12)$ $\cap(x\ge 2)\cap((x\ge 2)\cup(-8x+y\le 15))\cap((-8x+y\le 15)\cup(-x+3y\le 24))$ $\cap((-x+3y\le 24)\cup(y\le 12))\cap(y\le 12)\}$
C	$\{(x,y)\in\mathbb{R}^2_+ \mid (x\ge 2)\cap((x\ge 2)\cup(-4x+y\le 5))\cap(-4x+y>5)\cap(-x+y>4)$ $\cap(-x+2y>13)\cap((-x+2y\le 13)\cup(-x+5y\le 46))\cap((-x+5y\le 46)\cup(y\le 12))$ $\cap(y\le 12)\}$
D	$\{(x,y)\in\mathbb{R}^2_+ \mid (x\le 2)\cup(-6x+y\ge -9)\cup(-3x+11y\ge 90)\cup(y\ge 12))$ $\cap(x\ge 2)\cap((x\ge 2)\cup(-6x+y\le -9))\cap((-6x+y\le -9)\cup(-3x+11y\le 90))$ $\cap((-3x+11y\le 90)\cup(y\le 12))\cap(y\le 12)\}$
E	$\{(x,y)\in\mathbb{R}^2_+ \mid (x\le 2)\cup(-6x+y\ge -9)\cup(-x+3y\ge 24)\cup(-x+9y\ge 90)\cup(y\ge 12)$ $\cap(x\ge 2)\cap((x\ge 2)\cup(-6x+y\le -9))\cap((-6x+y\le -9)\cup(-x+3y\le 24))$ $\cap((-x+3y\le 24)\cup(-x+9y\le 90))\cap((-x+9y\le 90)\cup(y\le 12))\cap(y\le 12)\}$
$G-O$	$\{(x,y)\in\mathbb{R}^2_+ \mid (x>2)\cap(y<12)\cap(-6x+y<9)\cap(-x+3y<24)\cap(-x+5y<46)\}$

Table 5.5: Conditional Stability Regions for all DMUs

also to all models which postulate polyhedral production possibility sets, e.g. *Additive DEA models* (Charnes et al. 1985), *Assurance-Region models* (Thompson et al. 1986) and *polyhedral Cone-Ratio models* (Charnes et al. 1990).

The identification of CSRs is a computationally intensive problem. Motivated by this we described a procedure that has the potential for considerably reducing the significant computational burden. From existing research in the field, we know that explicitly identifying production possibility sets is possible for datasets with no more than 100-150 extreme-efficient units (see e.g. Olesen and Petersen 2003). In practice, the set of extreme-efficient units is often a small subset of the complete set of DMUs and this allows for explicitly characterising production possibility sets generated by fairly large datasets. However, computing CSRs requires additional computations, as it is based on the consecutive identification of production possibility sets. Our improved computational framework aims to reduce this additional computational load. Nevertheless, the applicability of CSA to larger datasets remains an open practical question. What is required to address this problem are computational experiments with datasets of various sizes. At the moment, we believe that for problems where CSA can be applied, the obvious advantages of such an application should make the additional computational effort very worthwhile. Finally, it is hoped that this research will encourage subsequent study into the subject, which will extend the theoretical framework presented here and, along with advances in computing capabilities, enhance the applicability of CSA to increasingly larger datasets.

Chapter 6

Cross-Evaluation: existing approaches

6.1 Introduction

The standard DEA approach for evaluating the efficiency of a set of units, allows each one to choose its own weighting scheme, such that its efficiency score is maximised (subject to some constraints). The efficiency scores calculated by this process of *self-appraisal* are also referred to as *simple-efficiencies*. Sexton et al. (1986), proposed to use the weighting scheme selected by a particular DMU to evaluate all of its peers. Efficiency scores resulting from this *peer-appraisal* process are given the name *cross-efficiencies*. The term *cross-evaluation* has also been introduced in the related literature (Doyle and Green 1995) to refer to the process of calculating cross-efficiency scores and the methodology behind this in general.

The most appealing property of cross-evaluation is that it improves the decision maker's discriminatory ability in evaluating different decisions/alternatives (in general) with respect to different criteria. To date, the use of cross-evaluation has spread to a number of different areas. In the *Multiple Criteria Decision Making (MCDM)* literature, a number of articles have suggested and explored the use of cross-evaluation as a tool for improving the discrimination among discrete *alternatives* (see e.g. Doyle 1995; Green and Doyle 1995; Bouyssou 1999; Sarkis 2000; Mavrotas and

Trifillis 2006). The use of cross-evaluation has also been suggested for the ranking of candidates in preferential elections (Green et al. 1996), and also in project and technology selection problems (Oral et al. 1991; Green et al. 1996; Shang and Shueyoshi 1995). Recently, Gregoriou et al. (2005, 2005b) reported the use of cross-evaluation, as part of a DEA-based methodology, to appraise the performance of *hedge-fund* and *Commodity Trading Advisors* (CTAs). Other applications of cross-evaluation, in the energy and manufacturing sector are reported/explored in Chen (2002), Sarkis and Talluri (2004) and Ertay and Ruan (2005).

Cross-evaluation could also be potentially appealing to central organisations in charge of funding DMUs. In the UK, the government's Department for Children, Schools and Families (earstwhile Department for Education and Skills) is considering the use of cross-evaluation as part of their methodology for the performance measurement of schools. It is planned that results of such a benchmarking exercise will be placed on a website, with the hope to assist the dissemination of best practice between schools (DfES 2005).

Despite the appeal of cross-evaluation for use in applications, its theoretical framework is a relatively understudied area, limited, to the best of our knowledge, to a small number of initial theoretical contributions (see e.g. Sexton et al. 1986; Doyle and Green 1995) and recent developments (Anderson et al. 2002). In this chapter we carry out a critical investigation of the theoretical framework for cross-evaluation and the computational tools used by traditional cross-evaluation approaches.

The outline of the chapter is as follows. In section 6.2 we introduce cross-evaluation and discuss its uses. In section 6.3 we explore the relationship of cross-evaluation to DEA production possibility set and discuss some interesting problems in providing a theoretical framework that integrates the two concepts. Finally, we review existing approaches for cross-evaluation in section 6.4, where we discuss in detail many flaws that we have discovered in their models.

6.2 Cross-Efficiency

6.2.1 Definition

Consider the non-linear form of the CCR model in (1.10), where DMU o is allowed to select weights to maximise its efficiency such that these weights do not yield efficiency scores greater than unity when applied to any DMU (including DMU o). In other words, all cross-efficiency scores are constrained to be less than or equal to one. A formal definition of cross-efficiency is as follows.

Definition 20 *Let (u_k^*, v_k^*) denote a set of optimal weights for DMU $k \in J$, obtained by solving a particular DEA model. The cross-efficiency of DMU $j \in J$ relative to DMU k, denoted h_{jk}, is given by:*

$$h_{jk} = \frac{u_k^* \mathbf{y}_j}{v_k^* \mathbf{x}_j} = \frac{O_{jk}}{I_{jk}} \tag{6.1}$$

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

After calculating the cross-efficiencies for all possible combinations of DMUs, the obtained information is summarised in an $n \times n$ *Cross-Efficiency Matrix* like the one given in table 6.1. Clearly, the entries of the leading diagonal are the simple efficiency scores. For simplicity we will also denote the simple-efficiencies by h_j instead of h_{jj}. Averaging across the rows[1] of the cross-efficiency matrix we obtain the average appraisal that DMU k assigns to its peers (denoted A_k), in other words the average cross-efficiency at the weights selected by DMU k. If the averaging is done across columns instead, we obtain the average appraisal of DMU k by its peers (denoted $\bar{h}_j$). Traditional cross-evaluation approaches have mostly focused on the second measure, taken to be the arithmetic mean of cross-efficiency scores, which is also referred to as average cross-efficiency. Doyle and Green (1994) also consider the median and the variance of cross-efficiency scores. Depending on whether we want to include the self-appraisal in the averaging, this might be done with or without the entries of the leading diagonal. We choose to include it and define $\bar{h}_j$ as follows

$$\bar{h}_j = \frac{1}{n} \sum_{k \in J} h_{jk} \tag{6.2}$$

[1] Or columns, depending on how the matrix is structured.

Rated DMUs	Rating DMUs					h_j
	1	2	3	...	n	
1	h_{11}	h_{12}	h_{13}	...	h_{1n}	h_1
2	h_{21}	h_{22}	h_{23}	...	h_{2n}	$\bar{h}_2$
3	h_{31}	h_{32}	h_{33}	...	h_{3n}	$\bar{h}_3$
...	...	...	...	...	...	...
n	h_{n1}	h_{n2}	h_{n3}	...	h_{nn}	$\bar{h}_n$
A_k	A_1	A_2	A_3	...	A_n	

Table 6.1: A Cross-Efficiency Matrix

To illustrate cross evaluation, we will use a small example of three DMUs that use two inputs and produce one output. The dataset is given in table 6.2. The production possibility set for one unit of output is given in figure 6.1. It is easy to notice that all three DMUs are extreme-efficient and hence possess multiple optimal weighting schemes. We discuss the implications of this in latter sections. For now, let us assume that DMU A has selected $w_1 = (1, 0, 3)$ as an optimal weighting scheme and DMUs B and C have selected $w_2 = (3, 4, 40)$. In figure 6.1, w_1 defines the hyperplane which supports the unbounded facet from A, and w_2 defines the hyperplane that supports facet BC.

The cross-efficiency of a particular DMU with respect to w_1 equates to the required radial reduction in its input levels, for that DMU to reach the hyperplane defined by w_1. For example, the cross efficiency of DMU B at the weights selected by DMU A (h_{AB}) is given by the radial reduction in B's inputs so that B is projected on w_1 at reference point B_1. Equivalently we can calculate h_{AB} by taking the ratio $\frac{|OB_1|}{|OB|}$. In a similar fashion we can use the other two reference points, A_2 and C_1, and calculate the remaining cross-efficiencies as follows: $h_{AC} = \frac{|OC_1|}{|OC|}$, $h_{BA} = h_{CA} = \frac{|OA_2|}{|OA|}$, $h_{BC} = h_{CB} = 1$.

Finally, we summarise the information on all calculated cross-efficiencies in the cross-efficiency matrix given in table 6.3. For example, the column corresponding to DMU A provides the cross-efficiencies assigned by DMU A's selected weights to all DMUs in the dataset, the corresponding reference points for these are A, B_1 and C_1 respectively. By averaging the entries in this column we obtain that the average appraisal by DMU A is $A_A = 0.7083$. In contrast, the row corresponding to DMU A provides the cross-efficiencies assigned to DMU A by DMU's B and C (and itself). Averaging the row entries we obtain that DMU A's average cross-efficiency is $\bar{h}_A = 0.8755$. Note that the columns for DMUs B and C are identical, as these DMUs have selected the same weighting

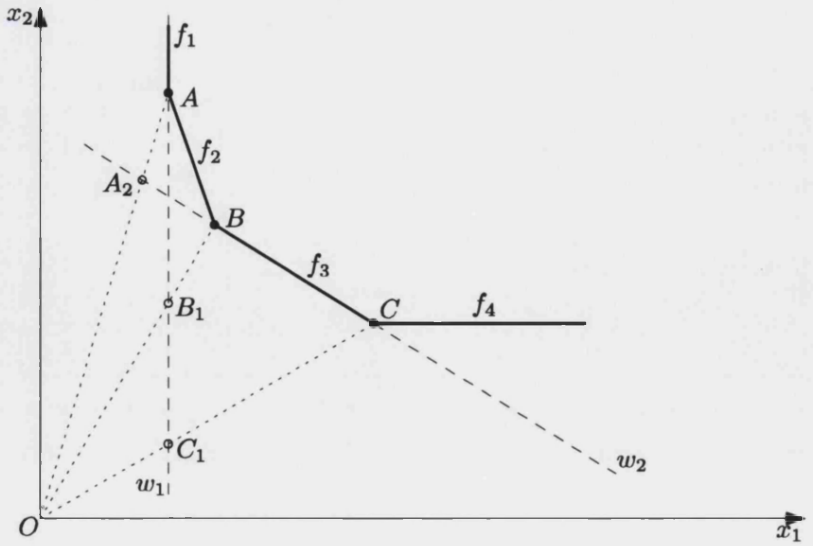


Figure 6.1: An illustration of cross-evaluation

scheme.

DMU	Output (y)	Input 1 (x_1)	Input 2 (x_2)
A	1	3	10
B	1	4	7
C	1	8	4

Table 6.2: Example Dataset

6.2.2 Uses of cross-evaluation

A first advantage of the average cross-efficiency measure over simple-efficiency is that the former can be used to rank all DMUs. In applications of DEA it might be desirable for the decision-maker/evaluator to construct a ranking of the units under evaluation based on their efficiency results. Using the simple-efficiency scores all inefficient units can be ranked, but not efficient ones, as they all obtain equal efficiency scores. Thus, no complete ranking of DMUs can be achieved by considering their simple-efficiency scores. By using the average cross-efficiency scores for all DMUs this problem is alleviated. In fact, although in theory an average cross-efficiency score of one (or 100%) is possible, in practice this would be a very unusual case requiring a 'strange' dataset. In general the possibility of ties among average cross-efficiency scores is very low and thus a unique

Rated DMUs	Rating DMUs			h_j
	A	B	C	
A	1	0.8163	0.8163	0.8775
B	0.7500	1	1	0.9167
C	0.3750	1	1	0.7917
A_k	0.7083	0.9388	0.9388	

Table 6.3: A Cross-Efficiency Matrix

ranking of all DMUs can be determined.

The super-efficiency model by Andersen and Petersen (1993) in which efficiency scores are not bounded by unity can also be used to rank all DMUs. This model has been very popular for determining such rankings (for a review of ranking methods in DEA, refer to Adler et al. 2002 and Angulo-Meza and Lins 2002). As mentioned in section 1.3.3, Banker and Chang (2006) conducted simulations to evaluate the performance of the super-efficiency model and report that this performs very poorly in ranking all DMUs.

Another important use of cross-evaluation is to reduce the effect of unrealistic, or *maverick*, weighting schemes used by some DMUs. In DEA, DMUs are allowed complete weight flexibility, i.e. they are free to choose the weights with which their efficiency will be evaluated. This is an important advantage of DEA to methods that employ 'externally' set weighting schemes which rely on the expertise of the evaluator, although it gives rise to the problem of some DMUs selecting weighting schemes that are 'unrealistic' under some managerial criteria. The case of zero weights is perhaps the most striking special case of this behaviour, which appears when DMUs only assign non-zero weights to the inputs and outputs that they make efficient use of, and leads to DMUs obtaining highly unrealistic efficiency scores. As discussed thoroughly in chapter 3, the selection of zero weights is of special theoretical importance to DEA, given its link with weak-efficiency. We would like to stress however that weighting schemes with all multipliers positive could also be unrealistic.

The most widely used methods for preventing DMUs from selecting unrealistic weighting schemes involve the introduction of additional constraints on the input and output weights in DEA models. The variety of methods for doing so are collectively referred to as *weight-restriction* methods and the additional constraints as *value-judgements* (see e.g. Thompson et al. 1986; Dyson and Thanassoulis 1988; Charnes et al. 1990). However, making value-judgements often

requires a priori information and relies on the expert knowledge of the modeller. In addition, value-judgements "also have something of the arbitrariness (authoritarianism) that exponents of DEA may have wished to escape" (Doyle and Green 1994). An alternative to weight-restricitons is cross-evaluation which reduces the effect of unrealistic weighting schemes since cross-efficiency values are a result of all available weighting schemes. In effect, "rather than have external weight restrictions applied by an expert, the dataset serves as the arbiter of good judgement by, in essence, creating its own weight restrictions" (Anderson et al. 2002).

Doyle and Green (1994) introduced the *maverick index* to identify DMUs that obtain unrealistic efficiency scores, which are also referred to as maverick DMUs. This index is based on average cross-efficiency, more specifically it measures the relative increment when shifting from average cross-efficiency to simple efficiency and is defined as follows:

$$M_j = \frac{h_j - \bar{h}_j}{\bar{h}_j} \tag{6.3}$$

Note that $M_j \geq 0$ because simple-efficiency is always greater than, or equal to average cross-efficiency. The rationale behind this is that a relatively high value of M_j suggests that, compared to weighting schemes selected by other DMUs, the weighting scheme selected by DMU j assigns a significantly higher efficiency score to itself on average, and therefore it should be viewed as an unrealistic weighting scheme. In addition, we can use the maverick index to identify all-round performers. These will be DMUs with relatively low maverick M_j values (see also Chen 2002 and Ertay and Ruan 2005 for applications of the maverick index).

6.3 A critical appraisal of the current theoretical framework

To date, the relationship between the theoretical framework for cross-evaluation and the attributes of DEA production possibility sets has not been explored in detail. In this section we undertake this task for both constant and variable returns-to-scale models and throw light on some serious problems in providing a consistent theoretical framework that integrates cross-evaluation with DEA production possibility sets.

Consider the illustration of cross-evaluation in the two-input one-output CRS case shown in

figure 6.2. The efficient frontier is shown in bold. The DEA production possibility set for the production of one unit of output can be defined as the intersection of five halfspaces corresponding to the hyperplanes denoted f_i (for $i = 1$ to 5). In turn, every hyperplane is associated with a unique weighting scheme given by the normal to the hyperplane. As we described earlier, the different cross-efficiency scores of a DMU with respect to these weighting schemes correspond to the different radial projections of DMUs on each of these hyperplanes. For example, the cross-efficiency of DMU E at weighting schemes 1 and 2 is given by the relative amount of radial reduction in the input levels of DMU E needed for E to reach points E_1 and E_2 on hyperplanes f_1 and f_2 respectively, which can be calculated by the ratios $\frac{OE_1}{OE}$ and $\frac{OE_2}{OE}$.

We can easily observe that DMUs A, B, C and D are DEA efficient, each with respect to 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 scheme for DMUs creates problems for cross-evaluation, which we explore in the next section. Another important problem we wish to highlight here is fact that reference points E_1, E_2, E_4 and E_5, used to define the cross-efficiencies of DMU E, lie outside the production possibility set.

In general, since $h_j \geq h_{jk}$, 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 appears in methodologies such as the *constrained facet analysis* (Bessent et al. 1988; Lang et al. 1995) and the *extended facet approach* (Olesen and Petersen 1996) that seek to control the selection of weighting schemes that contain zero multipliers by controlling the projection of DMUs on the frontier. More specifically these work by replacing 'unacceptable' projections (for instance a projection on a weakly-efficient facet) of inefficient DMUs on particular facets with the closest 'acceptable' ones on alternative facets and use the latter to specify new targets and calculate efficiency scores for DMUs. These new targets are not a part of the original production possibility set but of an enlarged one that was created by extrapolating existing facets.

In our case, a similar argument is deemed unreasonable; including all reference points for cross-evaluations in the production possibility set would lead to a highly unrealistic enlargement. For example, in figure 6.2 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. Taking this further, 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.

The above imply that cross-evaluation is based on a significant departure from fundamental assumptions of DEA. At this point we shall attempt to restore its credentials by reflecting further on what is really at stake. Cross-evaluation is not intended for setting targets for DMUs, but rather with measuring efficiency with the use of scoring 'recipes' used by peers. Therefore, the many reference points used in the calculation of cross-efficiencies should not be interpreted as targets and hence need not be included in the production possibility set. We have seen in chapter 2 that the efficient frontier can be identified by a collection of hyperplanes that are defined by different sets of weights. Taking each hyperplane alone, defines an alternative frontier that can be used to evaluate efficiencies. Consequently, instead of using a unique frontier to measure efficiency, cross-evaluation uses different frontiers and reports more than one efficiency scores which are 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. Hence, a fundamental weakness of existing approaches is that these 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 misleading interpretations. We illustrate these anomalies in the following section and present an alternative approach in the chapter 7.

Next we look at the VRS case which is unfortunately more problematic. Consider the one-input one-output example in Figure 6.3. The efficient frontier is given by the piecewise linear segment ABCD, adjoined by the rays AA′ and DD′ in the case of weak-efficiency. Input oriented cross-efficiencies for DMU E are calculated with respect to points E_1 and E_2 whereas E_3 and E_4 correspond to output oriented cross-evaluations (not all cross-evaluations are plotted). As before, E_2 and E_4 lie outside of the VRS production possibility set and clearly violate the convexity constraint of the BCC model. One additional problem is that input oriented cross-evaluations on some hyperplanes[2] result in negative cross-efficiency scores, as for example, the cross-efficiency of E with respect to point E_2 which lies in the negative orthant. Another problem is that some cross-evaluations for a particular DMU might correspond to weights that are not always compatible with

[2] To be more specific this only occurs decreasing returns-to-scale hyperplanes, i.e. the ones with a negative offset (β).

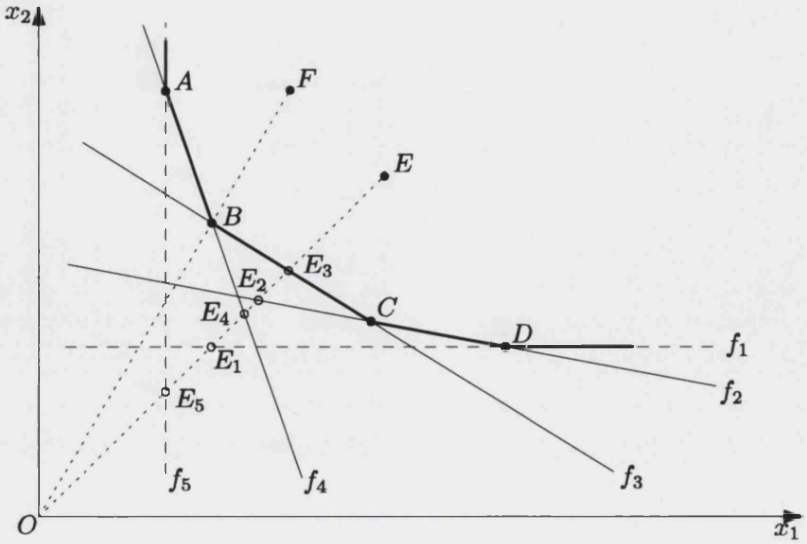


Figure 6.2: Cross-Evaluation and the CRS production possibility set

the returns-to-scale type exhibited by that DMU, as for example the input or output-oriented cross evaluation of DMU A, in figure 6.3, at the weighting scheme defined by the hyperplane supporting the facet BC. This contradicts the rationale behind the VRS model in which DMUs that exhibit increasing returns-to-scale are never compared to DMUs that exhibit decreasing returns-to-scale and vice-versa (see e.g. Tone 1996). As a result, the usefulness of cross-evaluation in VRS cases is reduced. To remedy this one could consider: a) using different (non-radial and non-oriented) efficiency measures for cross-evaluation and b) calculating the cross-efficiencies of a DMU only with respect to DMUs that exhibit compatible types of returns-to-scale. For the purposes of this study, we shall leave these issues open for further research and confine ourselves to the CRS case only.

6.4 A critical appraisal of traditional approaches

6.4.1 Implications of multiple optimality

In this section we examine problems that are specific to the traditional approaches for cross-evaluation. As we shall see these stem from the existence of multiple solutions for multiplier DEA

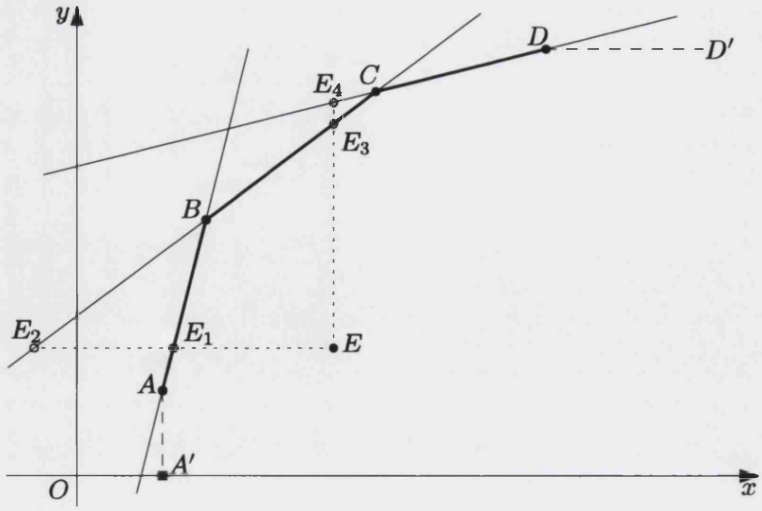


Figure 6.3: Cross-Evaluation and the VRS production possibility set

models, i.e. the existence of multiple sets of weighting schemes that yield the best efficiency score for a DMU. The implications for cross-evaluation are serious since cross-efficiency scores are entirely dependent on which of the multiple weighting schemes is chosen. In addition, the algorithms used to solve DEA models can only provide us with one of the multiple sets of weights.

To illustrate the implications of multiple optimality for cross-evaluation let us consider again the example given in figure 6.1. Each of the three DMUs A, B and C is efficient with respect to two sets of weights, but they can only select one of these. This implies many possible versions of the cross-efficiency matrix, of which we can only construct one. For this small example it is easy to work out the cross-efficiency matrices based on all possible scenarios. The results are summarised in table 6.4, where w_i refers to the weighting scheme associated with facet f_i, for $i = 1, ..., 4$. It is easy to observe that the average cross-efficiencies vary substantially across the eight scenarios. Overall, the maximum possible differences in average cross-efficiencies across all eight scenarions are: 20% for $\bar{h}_A$, 22.62% for $\bar{h}_B$ and 20.84% for $\bar{h}_C$.

Overall, selecting different weighting schemes can produce substantially different cross-efficiency scores and this renders the results unreliable. Traditionally, this problem has been tackled by the introduction of a secondary criterion, in the form of an objective function, in order to identify a unique set of weights for every DMU. More specifically, existing approaches select their weights by

Scenario	Weights			h_A	h_B	h_C
	A	B	C			
1	w_1	w_3	w_3	0.8776	0.9167	0.7917
2	w_2	w_2	w_3	0.9388	1.0000	0.7857
3	w_2	w_3	w_3	0.8776	1.0000	0.8929
4	w_2	w_2	w_4	0.8000	0.8571	0.7857
5	w_1	w_2	w_3	0.9388	0.9167	0.6845
6	w_1	w_2	w_4	0.8000	0.7738	0.6845
7	w_1	w_3	w_4	0.7388	0.7738	0.7917
8	w_2	w_3	w_4	0.7388	0.8571	0.8929

Table 6.4: Scenarios for the example in figure 6.1

identifying weighting schemes that either minimise or maximise the average cross-efficiency of a DMU's peers. Accordingly, they distinguish between aggressive and benevolent formulations. We shall highlight some inadequacies of these traditional approaches by identifying problems that are specific to particular models.

6.4.2 Existing models

We examine the aggressive formulations of four cross-evaluation approaches, taken from Doyle and Green (1995). All four approaches are implemented by two-stage models. The first stage maximises simple efficiency and the second stage introduces the secondary goal of minimising the cross-efficiencies of all other DMUs in some way, while of course keeping self-appraisal equal to that obtained in the first stage. This is summarised in (6.4).

$$\begin{aligned} &\text{Primary goal } \max\ h_k \\ &\text{Secondary goal } \min\ \frac{1}{n-1}\sum_{j\neq k}\frac{O_{jk}}{I_{jk}} \end{aligned} \tag{6.4}$$

The secondary goal in the above model involves the minimisation of the average appraisal that DMU k assigns to its peers. Clearly, this goal is non-linear, but even if this is ignored, there are problems concerning the choice among multiple optimal weights. This two stage approach has been suggested as a remedy for multiple optima in the first stage. It is not recognised however, that there may exist multiple optimal solutions even at the second stage, in which case the problem

of having to choose one among many remains. As mentioned earlier, such a choice has important implications because each weighting scheme will produce different cross-efficiency scores and hence different average cross-efficiency scores as well. As a result, choosing among weights has a knock-on effect on the ranking of DMUs. Nevertheless, it is reasonable to expect that the existence of multiple optimal weighting schemes would not be as common for average peer appraisals as it is for self- appraisals. Hence, the main problem with this approach stems from the non-linearity of the secondary goal which makes the model computationally intractable.

To tackle this computational difficulty, Sexton et al. (1986) and Doyle and Green (1995) respectively introduced models (6.5) and (6.6) as surrogates for model (6.4).

$$\begin{aligned} &\max\ h_k \\ &\min \sum_{j\neq k}(O_{jk} - I_{jk}) \end{aligned} \tag{6.5}$$

$$\begin{aligned} &\max\ h_k \\ &\min \frac{\sum_{j\neq k} O_{jk}}{\sum_{j\neq k} I_{jk}} \end{aligned} \tag{6.6}$$

Doyle and Green report that (6.6) is a better surrogate and show that its secondary goal is equivalent to selecting DMU k's optimal weighting scheme that minimises the efficiency of a composite DMU which includes the 'economy' of all DMUs except DMU k. The composite DMU is denoted by C^k and its input-output bundle is denoted by $(\mathbf{x}^k, \mathbf{y}^k)$ and calculated as follows:

$$(\mathbf{x}^k, \mathbf{y}^k) = (\sum_{j\neq k} \mathbf{x}_j, \sum_{j\neq k} \mathbf{y}_j) \tag{6.7}$$

The secondary goal in (6.6) is then replaced as shown below.

$$\frac{\sum_{j\neq k} O_{jk}}{\sum_{j\neq k} I_{jk}} = \frac{u_k \mathbf{y}^k}{v_k \mathbf{x}^k} \tag{6.8}$$

To illustrate how the surrogate model in (6.6) can select the wrong weighting schemes, we discuss two problematic cases. Similar examples can be constructed for (6.5). Table 6.5 provides

the data for the first case and table 6.6 for the second case. For both cases we shall examine the selection among optimal weights for DMU 1, i.e. $k = 1$. Finally, we denote the composite DMUs for the data in tables 6.5 and 6.6 respectively by $C_1^1 = (9,\ 36,\ 29)$ and $C_2^1 = (7,\ 28,\ 28)$.

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 6.5: 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 6.6: Data for case 2

To be able to evaluate the appropriateness of weight selection by the surrogate model we need to hold information on all of its optimal weighting schemes. These can be identified by using FMEL, for instance, as proposed in Appa and Williams (2006) and discussed in chapter 2. From this we can deduce that in both cases DMU 1 is efficient when using either of the two sets of weights given in Table 6.7. Note that the column A_t^1 denotes the average cross-efficiency of all DMUs except DMU 1 at weighting scheme t (for $t = 1,\ 2$).

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 6.7: Weighting schemes and their appraisals

DMU 1, being an 'aggressive' DMU, wants to minimise the average appraisal of its peers (A_t^1), so in the first case the correct choice would be weighting scheme two, since this minimises A_t^1. However, by implementing surrogate model (6.6), DMU 1 will choose weighting scheme one because this minimises the cross-efficiency of composite DMU C_1^1, thus clearly making a wrong choice.

We examine case two for the purpose of illustrating the possibility of multiple optimal weights in the second stage of the optimisation procedure. By implementing (6.6), the choice between the two weighting schemes would be arbitrary since they both assign the same efficiency score to composite DMU C_2^1. However, the true average appraisal of peers is clearly minimised at weight set two. Earlier, we mentioned that the choice among multiple optimal weights in (6.4) can affect the ranking of DMUs, and this example shows that this can be the case for surrogate models as well. In addition, the problem in this case is worse because the wrong choice can be made when selecting arbitrarily among multiple optimal weights for a DMU, as for example if we were to choose weighting scheme one for DMU A.

The final model that we discuss, by Doyle and Green (1995), is given below:

$$\begin{aligned} &\max\ h_{kk} \\ &\min \frac{O_{jk}}{I_{jk}} \ \forall\ k \neq j \end{aligned} \tag{6.9}$$

This model does not necessarily restrict DMUs to appraise all of their peers using the same weighting scheme. Instead, it allows DMUs to choose different weighting schemes for different cross-evaluations, such that all cross-efficiencies are separately minimised. One can argue against allowing the use of different weights in different cross-evaluations, on the basis of consistency in appraisal of peers. This rationale seems to contradict a statement by the same authors in an earlier article of theirs, 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 and Green 1994). Nevertheless, this approach does not face any problems like those described earlier for the previous approaches, and this would suggest that provided that the weights are consistently applied when calculating cross-efficiencies, the inclusion of a greater number of weighting schemes is a desirable

property for a cross-evaluation analysis.

Chapter 7

A new methodology for Cross-Evaluation

7.1 Introduction

In this chapter we introduce a novel approach for cross-evaluation (see also Appa, Argyris, and Williams 2006). This is based on the information obtained by an explicit identification of the DEA production possibility set. Our approach departs from the traditional rationale of peer appraisal within cross-evaluation and focuses on the evaluation of DMUs with respect to all available weighting-schemes that define facets of the production possibility set. We shall see later that this produces more meaningful cross-evaluation results. In addition, we introduce the concept of *under-achieving* DMUs and discuss methods for their identification, as well as the identification of unrealistic, or *maverick* weighting-schemes.

The outline of the chapter is as follows. In section 7.2 we discuss how the general philosophy of the existing approaches for cross-evaluation is problematic. We reveal that in certain cases this philosophy leads to results that contradict one of the main motivations for cross-evaluation, which is to reduce the effect of unrealistic weighting schemes. We address this issue, as well as the problems discussed in the previous chapter, by introducing a new approach for cross-evaluation in section 7.3. We provide a detailed discussion on how this approach provides more meaningful cross-

efficiency scores and describe how these can be used towards the better identification of maverick and under-achieving DMUs. In section 7.4 we provide an illustrative example for our new approach and close the chapter with some final remarks in section 7.6.

7.2 General problems

Here we highlight some problems regarding the general philosophy of traditional cross-evaluation approaches. The first point we would like to address is that aggressive formulations often result in the selection of maverick weighting schemes by DMUs. Since maverick weights are especially used by DMUs that specialise in a subset of inputs and outputs, it is very likely that such weights will provide the most conservative appraisal of peer DMUs. In the example in figure 6.2, DMU D is efficient at the weights associated with hyperplanes f_1 and f_2. It will choose the weighting scheme that minimises the appraisal of its peers i.e. the weighting scheme associated with f_1. However, f_1 supports a weakly-efficient facet and corresponds to a maverick weighting scheme which assigns a zero weight to input 1. An analogous observation holds for DMU A which will choose between f_5 rather than f_4. This type of behaviour contradicts one of the prime motivations for cross-evaluation, namely to reduce the effect of unrealistic weighting schemes. Note that this does not apply to benevolent surrogate models.

One way to resolve this issue is 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. Unfortunately, this comes with further complications, as shown below.

In general, imposing such weight restrictions is not guaranteed to be effective and might even result in assigning highly unrealistic efficiency scores for many 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 6.2. 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.

7.3 A New Approach for Cross-Evaluation

7.3.1 General description and advantages

Traditional approaches for cross-evaluation assume the unavailability of all weighting schemes and instead introduce secondary aggressive/benevolent criteria in order to choose one weighting scheme for each DMU. These approaches are inherently handicapped; although it might seem that these formulations are selected to provide an appropriate orientation to the analysis, this disguises the fact that a choice between these formulations is necessary if an analysis is to be carried out at all. In the real world it could be the case that neither aggressive nor benevolent formulations are suitable options for particular problems, or even that the choice between these might not be easy. For example, Green et al. (1996) report that in the case of ranking R&D projects, arguments can be made for the use of both options.

In our earlier discussion, we used all available weighting schemes for a particular DMU in order to check the validity of the results arrived at by the traditional approaches and, more importantly, identify correct solutions for all cases. In general, information on all sets of weights, obtained by the explicit identification of the production possibility set, can provide a solution for model (6.4) and thus eliminate the need for surrogate models. Motivated by the usefulness of this we propose a new approach for cross-evaluation based on the availability of explicit information on the production set (see also Appa, Argyris, and Williams (2006)).

More specifically, we utilise all of the weighting schemes in set $\bar{P}_C$. Any of the approaches described in chapter 2 can be used to explicitly identify this set. In particular we have used the FMEL method, building on the approach by Appa and Williams (2006) who evaluate a DMU's efficiency across all available weighting schemes as part of their new framework for the solution of DEA models, and we extend cross-evaluation to include all possible scoring combinations. By doing so, we need to depart from the standard definition of cross-efficiency (the efficiency of DMU j when using the weights selected by DMU k) which is now deemed insufficient. This is simply because many DMUs are efficient at more than one set of weights, and conversely because some weights might be optimal for more than one DMU, which would result in many repeated entries in the standard cross-efficiency matrix.

Formally, this gives rise to the following definition:

DMUs (j)	Weighting Scheme (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 7.1: The complete Cross-Efficiency Matrix

Definition 21 *The cross-efficiency of DMU* $j \in J$ *relative to weighting scheme* $(u_i, v_i) \in \bar{P}_C$, $i \in I_C = \{1, ..., t\}$, *denoted* c_{ji} *is given by:*

$$c_{ji} = \frac{u_i \mathbf{y}_j}{v_i \mathbf{x}_j} \tag{7.1}$$

The restructured, $n \times t$ - dimensional, cross-efficiency matrix containing all possible scoring combinations is given in table 7.1. Standard cross-evaluation procedures can only construct an incomplete cross-efficiency matrix, in the sense that it only contains information on a small subset of weighting schemes[1]. The following define the average cross efficiency and the maverick index for DMU j with respect to the complete cross-efficiency matrix:

$$\bar{c}_j = \frac{1}{t} \sum_i c_{ji} \tag{7.2}$$

$$m_j = \frac{h_j - \bar{c}_j}{\bar{c}_j} \tag{7.3}$$

Although this new approach seems structurally very similar to previous approaches, its philosophy is fundamentally different. Instead of averaging the efficiency appraisals by all peers we focus on 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) Cross-evaluation almost always achieves a unique ranking of DMUs through average cross-

[1] With an increase in the dimensions of the problem, the subset of weighting schemes obtained by standard cross-evaluation becomes increasingly smaller compared to the complete set of weighting schemes.

efficiency scores. For this reason, the amount of information on weights included in these scores is of utmost importance. Our approach is not restricted to establishing a unique weighting scheme for every DMU. By eliminating the need for aggressive or benevolent formulations we overcome *all* the problematic cases discussed earlier and therefore produce more meaningful average cross-efficiency scores, ranking of DMUs and maverick indices.

(2) Weight restrictions can now be introduced in cross-evaluation without further complications. More specifically, we can a priori identify unrealistic weighting schemes and exclude them from the analysis. Most importantly, since all weighting schemes are available, we need not worry about unrealistically evaluating some DMUs because of limited information. Hence, we can practically eliminate the effect of unrealistic weighting schemes on the average cross-efficiency scores.

(3) Unlike existing approaches, we are no longer restricted to input and output oriented efficiency measures. With all supporting hyperplanes 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; 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 earlier, for cross-evaluation under the VRS model.

7.3.2 Further extensions and tools

The new approach comes with a set of extensions and managerial tools which we describe here. We start by discussing how we can now identify maverick DMUs more effectively. The original maverick index in (6.3) compares a DMU's simple efficiency with its average cross-efficiency. The rationale is that if DMUs are using unrealistic weighting schemes the difference between these two efficiency measures will be high. However, this 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. To explain this further, consider the 2-input, 1-

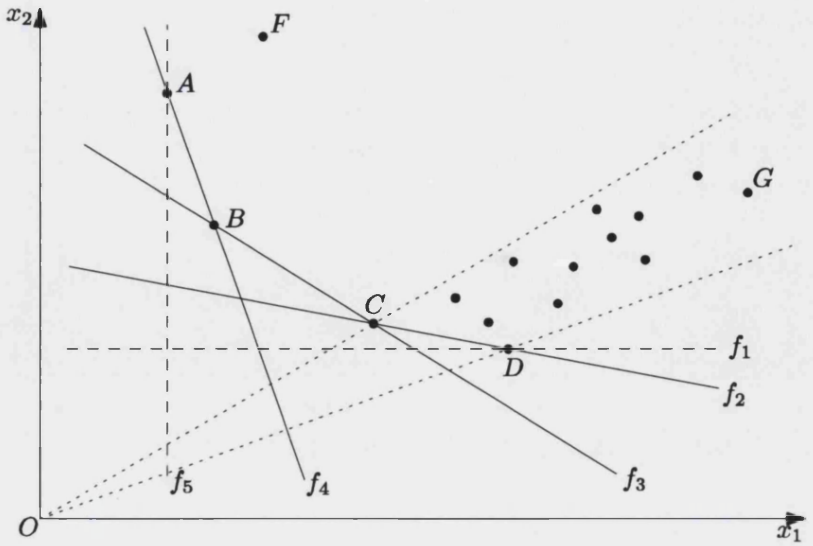


Figure 7.1: Identifying maverick and under-achieving DMUs

output CRS example in figure 7.1. By using m_j in (7.3) 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, in the sense that they would all choose one of D's optimal weighing schemes in their self-appraisals. 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:

$$W_j = \{(u_i, v_i) \in \bar{P}_C |\ c_{ji} = \max_i\{c_{ji}\}\} \tag{7.4}$$

Now consider the average efficiency of all DMUs on weighting scheme i:

$$\bar{f}_i = \frac{1}{n}\sum_j c_{ji} \tag{7.5}$$

For every DMU j we define:

$$\kappa_j = \max\{\bar{f}_i | (u_i, v_i) \in W_j\} \tag{7.6}$$

Finally, by utilising κ_j we define the new indicator as follows:

$$p_j = \frac{h_j - \kappa_j}{\kappa_j} \tag{7.7}$$

We have based the new indicator on κ_j. The rationale for this is that if a relatively high number of DMUs choose the same weighting scheme as DMU j, then we would not want to characterise DMU j as a maverick DMU. In such a case, κ_j would have a value relatively close to h_j, leading to a value for p_j relatively close to zero. In contrast, a low κ_j value relative to h_j will lead to a high value for p_j, pointing to maverick DMUs. For example in figure 7.1 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, which choose f_2, as mavericks. On the other hand, DMU A 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. We would like to note here that there are alternative ways to define κ_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.

The new indicator identifies unrealistic behaviour in a fundamentally different way from the conventional maverick index. Essentially, m_j examines the efficiency scores of a specific DMU on all weighting schemes and p_j examines the efficiency scores of all DMUs on a specific weighting scheme. One could use these indices in parallel in order to better identify maverick DMUs.

At this point we introduce the concept of *under-achieving* DMUs. It is important to differentiate between low-achieving and an under-achieving DMUs. The former is simply a DMU that achieves a low simple efficiency score whereas the latter is a DMU that is evaluated considerably lower than that of many other DMUs by its optimal weighting schemes. Hence, not only very inefficient DMUs can be under-achievers just as not only efficient DMUs can be mavericks. In figure 7.1 both DMUs F and G are low achievers but unlike F, G is also an under-achiever. We propose to identify under-achieving DMUs with use of their p_j indicator values. Notice that this can also take negative values (when $h_j < \kappa_j$) which is the case if the weighting scheme associated with κ_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.

Finally, we suggest that the motivation behind our approach is particularly relevant when trying to identify unrealistic weighting schemes and impose weight restrictions. The DEA literature contains a considerable amount of methods for imposing these (see Thanassoulis et al. 2004 for a review of methods for imposing weight restrictions in DEA). Most of these focus on the values of the input and output multipliers in each weighting scheme. The decision-maker usually has a view of how the production process taking place in DMUs should be represented in the values of the weights selected by DMUs. If a weighting scheme contains multipliers that are not in line with the decision maker's views, then that weighting scheme is termed unrealistic. Clearly, obtaining weighting schemes with realistic values for input and output weights is extremely important but in addition, as we illustrated in figure 7.1, it is often important to complement weight restriction analysis with information about how DMUs behave with respect to different weighting schemes. This issue has been relatively unexplored. We suggest that in addition to the current framework, assessing weighting schemes on the basis of their average efficiency appraisals 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 obtain 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.

7.4 Illustrative example

To illustrate the new approach we provide a simple 2-input, 1-output example using the hypothetical data for fifteen DMUs given in table 7.2. The production possibility set for this small example is given in figure 7.2. The efficient frontier is comprised of six facets of which the corresponding weighting schemes are given in table 7.3. We can easily observe that most DMUs make more efficient use of input 1 than input 2.

At this point we can consider whether some of the identified weighting schemes are unrealis-

tic. Two obvious candidates are weighting schemes 1 and 6 which include zero multipliers and correspond to weakly-efficient facets. Other possible candidates include weighting schemes 2 and 5 in which the weights for input 2 and input 1 respectively are very small. We suggested earlier that we should additionally consider the average efficiency appraisals by these weighting schemes (given in column $\bar{f}_i$ in table 7.3). Weighting scheme 2 has one of the highest average appraisals and weighting scheme 5 one of the lowest. This comes as no surprise, given the behaviour of DMUs in our dataset, namely the fact that most of them make more efficient use of input 1. With this in mind, it would be dubious to identify weighting scheme 2 as unrealistic. On the other hand we now have more evidence suggesting that weighting scheme 5 is unrealistic.

Having established which weighting schemes will be excluded we can proceed with the calculation of the cross-efficiency matrix. This is given in table 7.4. We shall examine three particular 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. Tables 7.5, 7.6 and 7.7 contain the results on simple and average cross-efficiencies, rankings, and the m_j and p_j values for all DMUs in these three cases respectively.

We start with comparing the results for m_j and p_j. In case one DMUs 1 and 7 achieve the two highest m_j values with m_1 slightly higher than m_7 which is no surprise given that they achieve their simple efficiency scores by using weighting schemes that contain zero multipliers. These values fail 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).

Moving to case two, DMUs 1 and 7 are still the most maverick DMUs in the context of index m_j but not for indicator p_j. The reason for this lies in 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

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	14	20	1
15	8	11	1

Table 7.2: Example dataset

only affected by one weighting scheme. Hence, following the exclusion of weighting schemes 1 and 6 only p_1 and p_7 decrease (calculated with weighting schemes 2 and 5 respectively) and all other p_j 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 achieves a lower p_j value than DMU 7 and DMU 6. This is because DMU 1 has a big family of peers with similar behaviour, whereas DMUs 7 and 6 are two of the relatively few DMUs which make more efficient use of input 2. For the same reason, DMU 5 achieves a relatively low value in index m_j but not in p_j.

In case three we additionally exclude weighting scheme 2. This forces DMUs 6 and 7 to achieve their simple efficiency with use of weighting scheme 4. As a result, 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.

Finally, consider DMU 14 which is clearly a low-achiever. Weighting scheme 3, selected by DMU 14 in its self appraisal, assigns considerably higher efficiencies to other DMUs than DMU 14. Hence, DMU 14 is also an under-achiever. This is clearly reflected in the relatively low value for p_{14} which remains the same throughout all above cases. No other p_j values merit the identification of other DMUs as under-achievers, in any of the three cases.

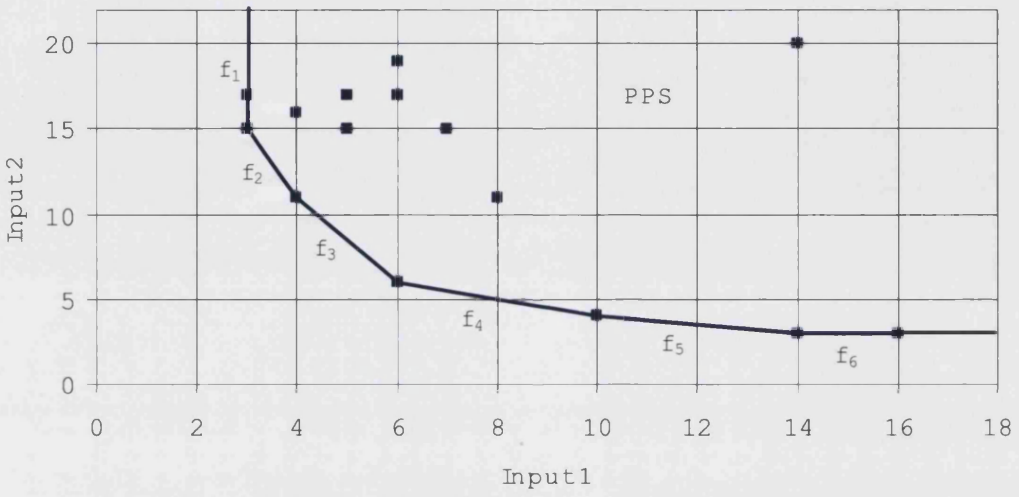


Figure 7.2: Example production possibility set

Weighting Scheme (i)	Weights			Efficient DMUs	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 7.3: The complete set of weighting schemes

DMUs	Weighting Schemes					
	f_1	f_2	f_3	f_4	f_5	f_6
1	1	0.9310	0.8571	0.4865	0.3662	0.1765
2	1	1	0.9333	0.5455	0.4127	0.2000
3	0.7500	1	1	0.6923	0.5417	0.2727
4	0.5000	0.9000	1	1	0.8667	0.5000
5	0.3000	0.6136	0.7241	1	1	0.7500
6	0.2143	0.4576	0.5526	0.9000	1	1
7	0.1875	0.4030	0.4884	0.8182	0.9286	1
8	07500	0.8438	0.8077	0.5000	0.3824	0.1875
9	0.6000	0.7714	0.7636	0.5143	0.4000	0.2000
10	0.4286	0.6269	0.6462	0.4865	0.3881	0.2000
11	0.6000	0.7297	0.7119	0.4615	0.3562	0.1765
12	0.5000	0.6585	0.6562	0.4500	0.3514	0.1765
13	0.5000	0.6279	0.6176	0.4091	0.3171	0.1579
14	0.2143	0.3553	0.3818	0.3333	0.2766	0.1500
15	0.3750	0.6279	0.6774	0.6000	0.5000	0.2727

Table 7.4: Example Cross-Efficiency Matrix

Case 1					
DMU	h_j	$\bar{c}_j$	rank	m_j	p_j
1	1	0.6362	7	0.5718	0.8563
2	1	0.6819	5	0.4665	0.3990
3	1	0.7095	3	0.4095	0.3668
4	1	0.7945	1	0.2587	0.3668
5	1	0.7313	2	0.3675	0.6207
6	1	0.6874	4	0.4547	0.8472
7	1	0.6376	6	0.5683	1.7633
8	0.8438	0.5786	8	0.4584	0.1805
9	0.7714	0.5416	9	0.4224	0.0792
10	0.6462	0.4629	13	0.3690	−0.1168
11	0.7297	0.5060	11	0.4422	0.0209
12	0.6585	0.4654	12	0.4148	−0.0787
13	0.6279	0.4383	14	0.4327	−0.1215
14	0.3818	0.2852	15	0.3386	−0.4782
15	0.6774	0.5088	10	0.3313	−0.0741

Table 7.5: Results, Case 1

Case 2					
DMU	h_j	$\bar{c}_j$	rank	m_j	p_j
1	0.931	0.6602	6	0.4102	0.7283
2	1	0.7229	5	0.3834	0.3990
3	1	0.8085	3	0.2369	0.3668
4	1	0.9417	1	0.0619	0.3668
5	1	0.8344	2	0.1984	0.6207
6	1	0.7276	4	0.3745	0.8472
7	0.9286	0.6596	7	0.4079	1.5660
8	0.8438	0.6335	8	0.3320	0.1805
9	0.7714	0.6123	9	0.2598	0.0792
10	0.6462	0.5372	12	0.2030	−0.1168
11	0.7297	0.5648	11	0.2919	0.0209
12	0.6585	0.5290	13	0.2447	−0.0787
13	0.6279	0.4929	14	0.2738	−0.1215
14	0.3818	0.3368	15	0.1338	−0.4782
15	0.6774	0.6013	10	0.1265	−0.0741

Table 7.6: Results, Case 2

Case 3					
DMU	h_j	$\bar{c}_j$	rank	m_j	p_j
1	0.9310	0.7582	5	0.2279	0.7283
2	1	0.8263	3	0.2103	0.3990
3	1	0.7984	2	0.1143	0.3668
4	1	0.9667	1	0.0345	0.3668
5	1	0.7792	4	0.2833	0.6207
6	0.9000	0.6367	8	0.4135	0.6624
7	0.8182	0.5699	13	0.4358	1.2610
8	0.8438	0.7172	6	0.1766	0.1805
9	0.7714	0.6831	7	0.1293	0.0792
10	0.6462	0.5869	12	0.1011	−0.1168
11	0.7297	0.6344	10	0.1503	0.0209
12	0.6585	0.5882	11	0.1195	−0.0787
13	0.6279	0.5515	14	0.1385	−0.1215
14	0.3818	0.3568	15	0.1071	−0.4782
15	0.6774	0.6351	9	0.0666	−0.0741

Table 7.7: Results, Case 3

7.5 Practical Implications

In this section we reflect on two practical implications of cross-evaluation that relate to the aggregation of average cross-efficiency scores and the structure of the cross-efficiency matrix.

The central tool in cross-evaluation is the average cross-efficiency score for every DMU which is obtained by taking the arithmetic mean of its cross-efficiency scores. The question then is why would one choose the arithmetic mean instead of the median, geometric mean or even a weighted average. In some cases the results obtained by choosing different aggregating options may vary substantially. Hence, one might argue that the use of the arithmetic mean (or even another option) is an arbitrary choice, albeit with potentially serious practical implications.

Although it is difficult to argue against the arbitrariness inherent in such aggregations, we feel that it is important to examine its practical implications in a relative context. Having obtained the information on the complete set of optimal weighting schemes and calculated the complete cross-efficiency matrix, the aggregation of these scores can proceed in a way that fits the scope of the analysis best. To start with, unrealistic weighting schemes can be omitted. Such weighting schemes can be identified immediately by inspecting the weights for zero values, whereas a more in-depth analysis can be performed by considering the average appraisals on different weighting schemes as suggested earlier. Overall this allows for completely removing the effect of unrealistic weighting schemes from the aggregate cross-efficiency scores. Following this, the information on all remaining weighting schemes can be useful in selecting a more realistic choice of aggregating cross-efficiency scores. More specifically, one can again consider the appraisals of DMUs on different weighting schemes, the choices of optimal weighting schemes by DMUs and the input-output multipliers across weighting schemes in combination with any prior expert knowledge for the problem, in order to choose a more meaningful way of aggregating the cross-efficiency scores, e.g. a particular weighted average. Overall then, although the choice of aggregation can always have serious practical implications, our argument is that at least by using the wealth of information included in the complete set of weights in combination with any expert knowledge one can make a more informed choice about how to aggregate cross-efficiency scores.

The second issue we would like to explore relates to the cross-efficiency matrix. The $n \times n$ structure of the traditional cross-efficiency matrix is something that might be appeal to practition-

ers who are interested in the idea of evaluating one DMU against another. However, with the use of all weighting schemes this structure is lost. Hence, there is scope in trying to combine the two distinct approaches in order to combine the advantages of the new approach with the appeal of the traditional cross-efficiency matrix.

One obvious way to do so would be to select one optimal weighting scheme per DMU according to pre-specified criteria as in the aggressive/benevolent formulations. As we explored in the previous chapter, such a choice comes with serious implications. Nevertheless, having recourse to the complete set of optimal weights allows for overcoming the computational issues of existing models and arriving at accurate results.

A more appealing approach would be to keep the structure of the traditional matrix but allow all weighting schemes (except perhaps the unrealistic ones) to influence the calculation of cross-efficiency scores. We illustrate this option in more detail below.

Let W_j be defined as in (7.4) and I_j be the index set for all weighting schemes in W_j. Further, define $t_j = |I_j|$. An alternative way of calculating average cross-efficiency scores would be as follows:

$$\hat{c}_j = \frac{1}{n}\left[\sum_{j\in J} \frac{\frac{1}{t_j}\left(\sum_{i\in I_j} u_i\right)\mathbf{y}_j}{\frac{1}{t_j}\left(\sum_{i\in I_j} v_i\right)\mathbf{x}_j}\right] = \frac{1}{n}\left[\sum_{j\in J} \frac{\sum_{i\in I_j} u_i\mathbf{y}_j}{\sum_{i\in I_j} v_i\mathbf{x}_j}\right] \tag{7.8}$$

As can be easily observed, this approach is based on aggregating the optimal weighting schemes for every DMU and applying these aggregate weights in calculating cross-efficiency scores. Note that, in the spirit of our earlier point, this aggregation could be made with use of a weighted average. Note also that this approach can be modified to include only weighting schemes that have not been identified as unrealistic.

This last option is a 'compromise' solution between traditional approaches for cross-evaluation and the new approach described earlier, in that it retains the $n \times n$ structure of the traditional cross-efficiency matrix which might be favoured by practitioners but at the same time calculates cross-efficiency scores in a way that allows information on all weighting schemes to take part in these calculations. There is no general answer as to which of the options for structuring the matrix is preferable. Instead, care should be taken to choose the one that suits the analysis most, so that

one can arrive at meaningful results.

7.6 Concluding remarks

Cross-evaluation is a cross-breed between standard DEA and efficiency evaluation by externally imposed criteria. By combining the two it inherits desirable attributes from both. On the one hand, by not allowing for total flexibility in self-appraisals, 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 chapter we identified some flaws in the general philosophy of existing approaches for cross-evaluation. Together with the specific computational problems discussed in the previous chapter, these render all existing methods highly problematic.

We addressed these problems by introducing, at least for the CRS case, a new approach for cross-evaluation which is based on computing and using the complete set of weights that define facets of the production possibility set. This overcomes all computational problems related to the surrogate models described in the previous chapter. Hence, it produces more meaningful results and allows for a better identification of maverick DMUs. We also introduced the concept of under-achieving DMUs and discussed how our framework can help in their identification. Most importantly, our approach overcomes the general problems of existing approaches discussed in this chapter. We are now able to safely introduce restrictions within cross-evaluation. This means that we can practically eliminate the impact of unrealistic weighting schemes on all cross-evaluation results. Finally, we offered a new way of assessing whether a weighting scheme is realistic, by suggesting that the average appraisal that this assigns to DMUs is considered. Our framework provides this information and can be used to construct empirical estimates of the acceptable ranges of input-output weights which, in the spirit of DEA, are generated only by information provided by the dataset itself. This can prove very useful in cases where there is no clear way of deciding on whether a particular weighting scheme is unrealistic by using existing weight-restriction approaches.

Bibliography

Adler, N., L. Friedman, and Z. Sinuany-Stern (2002). Review of ranking methods in the data envelopment analysis context. *European Journal of Operational Research 140*, 249–265.

Ali, A. and L. Seiford (1993). The mathematical programming approach to efficiency analysis. In H. Fried, C. Lovell, and S. Schmidt (Eds.), *The Measurement of Productive Efficiency*, Chapter 3, pp. 120–159. Oxford University Press.

Amirteimoori, A. and S. Kordrostami (2005). Efficient surfaces and an efficiency index in DEA: A constant returns to scale specification. *Applied Mathematics and Computation 163*, 683–691.

Andersen, P. and N. Petersen (1993). A procedure for ranking efficient units in data envelopment analysis. *Management Science 39*, 1261–1264.

Anderson, T., K. Hollingsworth, and L. Inman (2002). The fixed weighting nature of a cross-evaluation model. *Journal of Productivity Analysis 17*, 249–255.

Angulo-Meza, L. and M. Lins (2002). Review of methods for increasing discrimination in data envelopment analysis. *Annals of Operations Research 116*, 225–242.

Appa, G., N. Argyris, and S. Parthasarathy (2006). A faster and trimmer algorithm based on generators for solving DEA problems. Technical Report Working Paper LSEOR 06.84, Operational Research Group, Department of Management, London School of Economics and Political Science.

Appa, G., N. Argyris, and H. Williams (2006). A methodology for cross-evaluation in DEA. Technical Report Working Paper LSEOR 06.81, Operational Research Group, Department of Management, London School of Economics and Political Science.

Appa, G. and H. Williams (2006). A new framework for the solution of DEA models. *European Journal of Operational Research 172*(2), 604–615.

Banker, R. and H. Chang (2006). The super-efficiency procedure for outlier identification, not for ranking efficient units. *European Journal of Operational Research 175*, 1311–1320.

Banker, R., A. Charnes, and W. Cooper (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science 30*(9), 1078–1092.

Banker, R., S. Das, and S. Datar (1989). Analysis of cost variances for management control in hospitals. *Research in Governmental and Nonprofit Accounting 5*, 269–291.

Banker, R. and J. Gifford (1988). A relative efficiency model for the evaluation of public health nurce productivity. Technical report, Melon University Mimeo, Carnegie.

Bessent, A., W. Bessent, and J. Elam (1988). Efficiency frontier determination by constrained facet analysis. *Operations Research 36*(5), 785–796.

Boljuncic, V. (2006). Sensitivity analysis of an efficiennt DMU in DEA model with variable returns to scale (VRS). *Journal of Productivity Analysis 25*, 173–192.

Bouyssou, D. (1999). Using DEA as a tool for MCDM: Some remarks. *The Journal of the Operational Research Society 50*(9), 974–978.

Briec, W. and H. Leleu (2003). Dual representations of non-parametric technologies and measurement of technical efficiency. *Journal of Productivity Analysis 20*, 71–96.

Bron, C. and J. Kerbosch (1973). Algorithm 457-finding all cliques of an undirected graph. *Communications of ACM 16*, 575–583.

Chambers, R., Y. Chung, and R. Fare (1996). Benefit and distance functions. *Journal of Economic Theory 70*, 407–419.

Chambers, R., Y. Chung, and R. Fare (1998). Profit, directional distance functions and nerlovian efficiency. *Journal of Optimization Theory and Applications 98*(2), 351–364.

Charnes, A. and W. Cooper (1962). Programming with linear fractional functionals. *Naval Research Logistics Quarterly 9*, 181–185.

Charnes, A. and W. Cooper (1968). Structural sensitivity analysis in linear programming and an exact product form left inverse. *Naval Research Logistics Quarterly 15*, 233–253.

Charnes, A. and W. Cooper (1985). Preface to topics in data envelopment analysis. *Annals of Operations Research 2*, 59–94.

Charnes, A., W. Cooper, B.Golany, and L.Seiford (1985). Foundations of data envelopment analysis for pareto-koopmans efficient empirical production functions. *Journal of Econometrics 30*, 91–107.

Charnes, A., W. Cooper, Z. Huang, and D. Sun (1990). Polyhedral cone-ratio DEA models with with an illustrative application to large commercial banks. *Journal of Econometrics 46*, 73–91.

Charnes, A., W. Cooper, A. Lewin, R. Morey, and J. Rousseau (1985). Sensitivity and sability analysis in DEA. *Annals of Operations Research 2*, 139–150.

Charnes, A., W. Cooper, and E. Rhodes (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research 2*, 429–444.

Charnes, A., W. Cooper, and R. Thrall (1991). A structure for classifying and characterizing efficiency and inefficiency in data envelopment analysis. *Journal of Productivity Analysis 2*, 197–237.

Charnes, A., S. Haag, P. Jaska, and J. Semple (1992). Sensitivity of efficiency calculations in the additive model of data envelopment analysis. *Journal of Systems Science 23*, 789–798.

Charnes, A., J. Rousseau, and J. Semple (1996). Sensitivity and stability of efficiency classifications in DEA. *Journal of Productivity Analysis 7*, 5–18.

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.

Cooper, W., S. Li, L. Seiford, K. Tone, R. Thrall, and J. Zhu (2001). Sensitivity and stability analysis in DEA: Some recent developments. *Journal of Productivity Analysis 15*, 217–246.

Cooper, W., J. Ruiz, and I. Sirvent (2006). Choosing weights from alternative optimal solutions of dual multiplier models in DEA. *European Journal of Operational Research 180*, 443–458.

Cooper, W., L. Seiford, and J.Zhu (2004). Data envelopment analysis: History, models and interpretations. In W. Cooper, L. Seiford, and J. Zhu (Eds.), *Handbook on Data Envelopment Analysis*, pp. 1–39. Kluwer Academic Publishers.

Cooper, W., L. Shanling, L. Seiford, and J.Zhu (2004). *Sensitivity Analysis in DEA*, Chapter 3, pp. 75–97. Kluwer Academic Publishers.

Dantzig, G. and B. Eaves (1973). Fourier-motzkin elimination and its dual. *Journal of Combinatorial Theory 14*, 288–297.

Debreu, G. (1951). The coefficient of resource utilization. *Econometrica 19*(3), 273–292.

DfES (2005). Private communication.

Doyle, J. (1995). Multiattribute choice for the lazy decision maker: Let the alternatives decide. *Organizational Behavior and Human Decsision Processes 62*(1), 87–100.

Doyle, J. and R. Green (1994). Efficiency and cross-efficiency in DEA; derivations, menings and uses. *The Journal of The Operational Research Society 45*(5), 567–578.

Doyle, J. and R. Green (1995). Cross-evaluation in DEA: Improving discrimination among DMUs. *Information Systems and Operational Research 33*(3), 205–222.

Dula, J. (1997). *An Algorithm for DEA*. The University of Mississipi Paper No. MS 38677.

Dula, J. and B. Hickman (1997). Effects of excluding the column being scored from the DEA envelopment LP technology matrix. *Journao of the Operational Research Society 48*, 1001–1012.

Dyson, R. and E. Thanassoulis (1988). Reducing weight flexibility in data envelopment analysis. *Journal of the Operational Research Society 39*, 563–576.

Ertay, T. and D. Ruan (2005). Data envelopment analysis decision model for optimal operator allocation in CMS. *European Journal of Operational Research 164*, 800–810.

Fare, R., S. Grosskopf, and C. Lovell (1985). *The Measurement of Efficiency of Production.* Boston: Kluwer-Nijhoff Publishing.

Fare, R. and C. Lovell (1978). Measuring technical efficiency of production. *Journal of Economic Theory 19*(1), 150–162.

Farrell, M. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society 120*(3), 253–281.

Forsund, F. and L. Hjalmarsson (2004). Are all scales optimal in DEA? theory and empirical evidence. *Journal of Productivity Analysis 21*, 25–48.

Gonzalez-Lima, M., R. Tapia, and R. Thrall (1996). On the construction of strong complementarity slackness solutions for DEA linear programming problems using a primal-dual interior-point method. *Annals of Operations Research 66*, 139–162.

Green, R. and J. Doyle (1995). On maximising discrimination in multiple criteria decision making. *The Journal of the Operational Research Society 46*(2), 192–204.

Green, R., J. Doyle, and W. Cook (1996). Preference voting and project ranking using DEA and cross-evaluation. *European Journal of Operational Research 90*, 461–472.

Gregoriou, G., F. Rouah, S. Satchell, and F. Diz (2005). Simple and cross efficiency of CTAs using data envelopment analysis. *The European Journal of Finance 11*(5), 393–409.

Gregoriou, G., K. Sedzro, and J. Zhu (2005). Hedge fund performance appraisal using data envelopment analysis. *European Journal of Operational Research 164*, 555–571.

Huang, Z., S. Li, and J. Rousseau (1993). Determining rates of change in data envelopment analysis. Technical Report Research report CCS 713, Center of Cybernetic Studies, The university of Texas at Austin.

Huang, Z., S. Li, and J. Rousseau (1997). Determining rates of change in data envelopment analysis. *Journal of the Operational Research Society 48*, 591–599.

Jahanshahloo, G., F. Lofti, and M. Zohrebandian (2005). Finding the piecewise linear production function in data envelopment analysis. *Applied Mathematics and Computation 163*, 483–488.

Jahanshahloo, G. A. F. H., N. Shoza, M. Sanei, and G. Tohidi (2005a). Sensitivity and stability analysis in data envelopment analysis. *Journal of the Operational Research Society 56*, 342–345.

Jahanshahloo, G. A. F. H., N. Shoza, M. Sanei, and G. Tohidi (2005b). Sensitivity and stability analysis in DEA. *Applied Mathematics and Computation 169*, 897–904.

Koopmans, T. (1951). An analysis of production as an efficient combination of activities. In T. Koopmans (Ed.), *Activity Analysis of Production and Allocation.* New York: John Wiley and Sons, Inc.

Krivonozhko, V., O. Utkin, A. Volodin, and I. Stablin (2005). About the structure of boundary points in DEA. *Journal of The Operational Research Society 56*(12), 1373–1378.

Lang, P., O. Yolalan, and O.Kettani (1995). Controlled envelopment by face extension in DEA. *Journal of the Operational Research Society 46*(4), 473–491.

Lovell, C. (1993). Production frontiers and productive efficiency. In H. Fried, C. Lovell, and S. Schmidt (Eds.), *The Measurement of Productive Efficiency.* Oxford University Press.

Lovell, C. and A. Rouse (2003). Equivalent standard DEA models to provide superefficiency scores. *Journal of the Operational Research Society 54*, 101–108.

Mavrotas, G. and P. Trifillis (2006). Multicriteria decision analysis with minimum information: Combining DEA with MAVT. *Computers and Operations Research 33*, 2083–2098.

Motzkin, T., H. Raiffa, G.L.Thompson, and R.M.Thrall (1953). The double description method. In H. Kuhn and A. Tucker (Eds.), *Contribution to the Theory of Games, Vol. II*, pp. 81–103. Princeton University Press.

Nemhauser, G. and L. Wolsey (1988). *Integer and Combinatorial Optimization.* Wiley, New York.

Neralic, L. (1997). Sensitivity in data envelopment analysis for arbitrary perturbations of data. *Glasnic Matematicki Ser. III 32*, 315–335.

Neralic, L. and R. Wendell (2004). Sensitivity in data envelopment analysis using an approximate inverse matrix. *Journal of the Operational Research Society 55*, 1187–1193.

Olesen, O. and N. Petersen (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. and N. Petersen (2003). Identification and use of efficient faces and facets in DEA. *Journal of Productivity Analysis 20*, 323–360.

Oral, M., O. Kettani, and P. Lang (1991). A methodology for collective evaluation and selection of industrial RD projects. *Mnagement Science 37*(7), 871–885.

Pitakong, U., P. Brockett, J. Mote, and J. Rousseau (1998). Identification of pareto-efficient facets in data envelopment analysis. *European Journal of Opeartional Research 109*, 559–570.

Portela, M. and E. Thanassoulis (2006). Zero weights and non-zero slacks: Different solutions to the same problem. *Annals of Operations Research 145*, 129–147.

Raty, T. (2002). Efficient facet efficiency index: A variable returns to scale specification. *Journal of Productivity Analysis 17*, 65–82.

Rockafellar, T. (1970). *Convex Analysis.* Princeton University Press, New Jersey.

Russell, R. (1988). On the axiomatic approach to the measurement of technical efficiency. In W. Eichhorn (Ed.), *Measurement in Economics: Theory and Applications of Economic Indices*, pp. 207–217. Heidelberg: Physica Verlag.

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. and S. Talluri (2004). Performance based clustering for benchmarking of US airports. *Transportation Research Part A-Policy and Practice 38*(5), 329–346.

Schrijver, A. (2000). *Theory of Linear and Inreger Programming.* Wiley, Chichester UK.

Seiford, L. and J. Zhu (1998a). Sensitivity analysis of DEA models for simultaneous changes in all the data. *Journal of the Operational Research Society 49*, 106–1071.

Seiford, L. and J. Zhu (1998b). Stability regions for maintaining efficiency in data envelopment analysis. *European Journal of Operational Research 108*, 127–139.

Seiford, L. and J. Zhu (1999). Infeasibility of super-efficiency data envelopment analysis. *INFOR 37*, 174–187.

Sexton, T., R. Silkman, and A. Hogan (1986). Data envelopment analysis: Critique and extensions. In R. Silkman (Ed.), *Measurin Efficiency: An Assessment of Data Envelopment Analysis*, pp. 73–105. San Francisco: Jossey Bass.

Shang, J. and T. Shueyoshi (1995). A unified framework for the selection of a flexible manufacturing system. *European Journal of Operational Research 85*, 297–315.

Shepard, R. (1953). *Cost and Production Functions.* Princeton: Princeton University Press.

Shepard, R. (1970). *Theory of Cost and Production Functions.* Princeton: Princeton University Press.

Thanassoulis, E., M. Portella, and R. Allen (2004). Incorporating value judgements in DEA. In W. Cooper, L. M. Seiford, and J. Zhu (Eds.), *Handbook on Data Envelopment Analysis*, Chapter 4, pp. 99–138. Kluwer Academic Publishers.

Thompson, R., D. Dharmapala, M. Gonzalez-Lima, and R. Thrall (1996). DEA multiplier analytic center sensitivity with an illustrative application to oil companies. *Annals of Operations Research 66*, 163–177.

Thompson, R., F. S. Jr, R. Thrall, and B. Smith (1986). Comparative site evaluations for locating a high-energy physics lab in texas. *Interfaces 16*, 35–49.

Thomson, R., P. Dharmapala, and R. Thrall (1994). Sensitivity analysis of efficiency measures with application to kansas farming and illinois coal mining. In A. Charnes, W. Cooper, A. Lewin, and L. Seiford (Eds.), *Data Envelopment Analysis: Theory, Methodology and Applications*, pp. 393–422. Kluwer Academic Publishers.

Thrall, R. (1996). Duality, classification and slacks in DEA. *Annalls of Operations Research 66*, 109–138.

Tone, K. (1996). A simple characterization of returns to scale in DEA. *Journal of the Operations Research Society of Japan 39*, 604–613.

Williams, H. (1986). Fourier's method of linear programming and its dual. *The American Mathematical Monthly 93*, 681–695.

Yu, G., Q. Wei, and P. Brockett (1996). A generalized data envelopment analysis model: A unification and extension of existing methods for efficiency analysis of decision making units. *Annals of Operations Research 66*, 47–89.

Yu, G., Q. Wei, P. Brockett, and L. Zhou (1996). Construction of all DEA efficient surfaces of the production possibility set under the generalised data envelopment analysis model. *European Journal of Operational Research 95*, 491–510.

Zhu, J. (1996). Robustness of the efficient DMUs in data envelopment analysis. *European Journal of Operational Research 129*, 443–455.

Zhu, J. (2001). Super-efficiency and DEA sensitivity analysis. *European Journal of Operational Research 129*, 443–455.