Series on Advanced Economic Issues
Faculty of Economics, VŠB-TU Ostrava

Mehdi Toloo

DATA ENVELOPMENT ANALYSIS
WITH SELECTED MODELS AND APPLICATIONS

Ostrava, 2014

Mehdi Toloo
Department of Business Administration
Faculty of Economics
VŠB-Technical University of Ostrava,
Sokolská 33
701 21 Ostrava, CZ
mehdi.toloo@vsb.cz

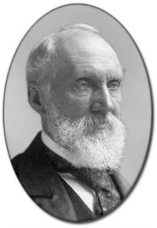
Reviews
Adel Hatami-Marbini, Université catholique de Louvain, Belgium
Sahand Daneshvar, Eastern Mediterranean University, Turkey
Reza Farzipour Saen, Islamic Azad University, Iran

This publication has been elaborated in the framework of the project *Support research and development in the Moravian-Silesian Region 2013 DT 1 – International research teams* (02613/2013/RRC). Financed from the budget of the Moravian-Silesian Region. It has been also supported by the Czech Science Foundation (GACR project 14-31593S) and through European Social Fund (OPVK project CZ.1.07/2.3.00/20.0296).

The text should be cited as follows: Toloo, M., (2014). *Data Envelopment Analysis with Selected Models and Applications*, SAEI, Vol. 30. Ostrava: VŠB-TU Ostrava.

Printed in Grafico, s.r.o.
Cover design by MD communications, s.r.o.
ISBN 978-80-248-3738-3

Preface



Lord William Thomson Kelvin
(1824-1907)

If you cannot measure it, you cannot improve it.

Lord William Thomson Kelvin

In microeconomics a production function is a mathematical function that transforms all combinations of inputs of an entity, firm or organization into the output. Given the set of all technically feasible combinations of outputs and inputs, only the combinations encompassing a maximum output for a specified set of inputs would constitute the production function. Data Envelopment Analysis (DEA), which has initially been originated by Charnes, Cooper and Rhodes in 1978, is a well-known non-parametric mathematical method with the aim of estimating the production function. In fact, DEA evaluates the relative performance of a set of homogeneous decision making units with multiple inputs and multiple outputs.

This book covers some basic DEA models and disregards more complicated ones, such as network DEA, and mainly stresses on the importance of weights in DEA and some of their applications. As a result, this book mostly considers the multiplier form of DEA models to extend some new approaches, however the envelopment forms are introduced in some possible approaches. This book also aims at dealing with some innovative uses of binary variables in extended DEA model formulations. The auxiliary variables enable us formulate Mixed Integer Programming (MIP) DEA models for addressing the problem of finding a single efficient and ranking efficient DMUs. In some cases, the status of input(s) or output(s) measure is unknown and binary variables are utilized to accommodate these flexible measures. Furthermore, the binary variables approach tackles the problem of selecting input or output measures.

The book also stresses the mathematical aspects of selected DEA models and their extensions so as to illustrate their potential uses with applications to different contexts, such as banking industry in the Czech Republic, financing decision problem, technology selection problem, facility layout design problem, and selecting the best tennis player. In addition, the majority of the extended models in this book can be extended to some other DEA models, such as slacks-based measures, hybrids, non-discretionary and fuzzy DEA which are applicable on some other contexts.

This research-based book contains six chapters as follows:

The first chapter (General Discussion) starts with a simple numerical example to explain the concept of relative efficiency and to clarify the importance of input and output weights in measuring the efficiency score. Then these basic concepts are extended to some more complex cases. Efficient frontiers and projection points are illustrated by means of some constructive and insightful graphs.

The second chapter (Basic DEA Models) presents both envelopment and multiplier forms of the DEA models in the presence of multiple inputs and multiple outputs. However, this book mainly focuses on multiplier form of DEA models. In addition, this chapter illustrates the role of each axiom to construct the production possibility set (PPS). It is also concerned with some DEA models to deal with pure input data as long as with pure output data set. Apart from basic input- and output-oriented DEA models with different returns to scale, the chapter includes a model that combines both orientations. Three different case studies involving banking industry, technology selection, and asset financing are provided in this section.

In chapter 3 (GAMS Software), we briefly introduce General Algebraic Modeling System (GAMS) software, a modeling system for linear, nonlinear and mixed integer optimization problems for solving DEA models.

Chapter 4 (Weights in DEA) treats the weights in DEA and their importance along with various weight restrictions and common set of weights (CSW) approaches. The chapter includes Assurance Region (AR) and Assurance Region Global (ARG) methods to restrict weight flexibility in DEA. Two DEA models with different types of efficiency, i.e. minsum and minimax, with their integrated versions are introduced in this chapter. The evaluation of facility layout design problem is addressed as a numerical example.

Chapter 5 (Best Efficient Unit) considers CSW and binary variable approaches as the main tool for developing models that have the capability to find the most efficient DMU and also rank DMUs. We cover WEI/WEO data sets along with multiple input and multiple output data set. Some epsilon-free DEA models are introduced to overcome the problem of finding a set of positive weights. The problem of finding the most cost efficient under uncertain input prices are also discussed. Two real data set involving professional tennis players and Turkish automotive company are rendered to validate the approaches in this chapter.

Chapter 6 (Data Selection in DEA) closes the book by considering data selection problem in DEA and presenting some modifications of the standard DEA models to accommodate flexible and selective measures. To deal with these problems, two multiplier and envelopment DEA models are developed where each model contains two alternative approaches: individual and integrated models. Individual approach classifies flexible measures and identifies selective measures for each DMU, and aggregate approach accommodates these measures using integrated

DEA models. We present three case studies to examine and validate the approaches in this chapter.

Evidently, my deepest gratitude and love go to my family, Laleh and Arad, for supporting me in writing this book. Ronak Azizi saved me a lot of trouble by tackling all formatting issues in Microsoft Word. Last, but certainly not least, I would like to extend my thanks to my friend, Dr. Adel Hatami-Marbini, for helping me with editing the book and invaluable ideas and comments.

This publication has been elaborated in the framework of the project *Support research and development in the Moravian-Silesian Region 2013 DT 1 – International research teams* (02613/2013/RRC). Financed from the budget of the Moravian-Silesian Region. It has been also supported by the Czech Science Foundation (GACR project 14-31593S) and through European Social Fund (OPVK project CZ.1.07/2.3.00/20.0296).

Mehdi Toloo, Ph.D.

Contents

List of Abbreviations

AHP	Analytic Hierarchy Process
AMT	Advanced Manufacturing Technology
ATP	Association of Tennis Professionals
BCC	Banker, Charnes, Cooper's model in 1984
CCR	Charnes, Cooper, Rhodes's model in 1978
CE	Cost Efficiency
CRS	Constant Returns to Scale
CSW	Common Set of Weights
DEA	Data Envelopment Analysis
DMU	Decision Making Unit
FLD	Facility Layout Design
FMS	Flexible Manufacturing System
GAMS	General Algebraic Modeling System, an optimization software
KKT	Karush-Kuhn-Tucker
LP	Linear Programming
MCDM	Multiple Criteria Decision Making
MIP	Mixed Integer Programming
MINLP	Mixed Integer Non-Linear Programming
VRS	Variable Returns to Scale
WCM	World Class Manufacturing

Glossary of symbols

v_i, d_j, θ	lowercases in italic denote scalars or variables
$\boldsymbol{v}, \boldsymbol{d}$	lowercases in boldface denote vectors
$\mathbf{X}, \mathbf{Y}$	uppercases in boldface denote matrices
(a, b)	open interval $a < x < b$
$[a, b]$	closed interval $a \leq x \leq b$
$\mathbf{0}_n$	origin in $\mathbb{R}^n$, i.e. $(0, \dots, 0) \in \mathbb{R}^n$
$\mathbf{1}_n$	$(1, \dots, 1) \in \mathbb{R}^n$
$\boldsymbol{e}_j$	j^{th} unit vector; i.e. $\boldsymbol{e}_j = (e_{ij})$, presents $e_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$
$\boldsymbol{v}^*, v_i^*$	Asterisk denotes the optimal value of a vector/variable

This Book is dedicated to my dear and loving wife, Laleh, and my curious son, Arad, who helped me with patience and kindness.

Thank you

CHAPTER 1

General Discussion

Limitation of resources is an undeniable fact, which is dealt with by many organizations such as business firms, banks, hospitals, universities, etc. Hence, improving the performance of resource utilization for organizations is one of the most important concerns of managers. As a result, a manager must evaluate frequently the performance of the organization. One essential step to improve the performance of organizations is to measure the efficiency. In general, the efficiency score of an organization with multiple inputs and multiple outputs is defined as the ratio of the weighted sum of outputs to the weighted sum of inputs. In this definition, the efficiency score is closely related to the determination of weights while different weights leads to different efficiency scores. Experts and managers usually, based on their experiments, assign the values to these weights. In contrast to this method, data envelopment analysis (DEA) finds the optimal weights with the aim of maximizing the efficiency score for each organization. These weights are obtained precisely from the data set (inputs and outputs) and may differ from one organization to another. In the rest of this chapter, the concept of input and output weights in evaluation performance is discussed.

1. 1 A simple case (single input and single output)

Suppose there are 12 hospitals that are labeled H_1 to H_{12} at the head of each row in Table 1–1. The second and third columns are the number of staff and the number of patients (measured in 100 persons/month), respectively. Let us start our discussion with comparing H_1 and H_5. Both hospitals have the same amount of staff, whereas more patients are admitted to H_5 compared to H_1; hence H_5 works better than H_1 and is relatively more efficient. In this case, we say that H_5 *dominates* H_1 (or equivalently H_1 is *dominated* by H_5). In a similar manner, consider H_8 and H_{12}: with the same quantity of admitted patients, the number of employed staff in the former hospital is less than the latter one which means H_8 dominates H_{12} and hence is relatively more efficient.

Table 1–1 Hospital case with single input and single output

Hospitals	Staff	Patients	$\frac{\text{Patients}}{\text{Staff}}$	e_j
H_1	173	222	1.283	0.659
H_2	280	240	0.857	0.440
H_3	225	212	0.942	0.484
H_4	246	297	1.207	0.620
H_5	173	248	1.434	0.736
H_6	323	316	0.978	0.502
H_7	253	342	1.352	0.694
H_8	212	362	1.708	0.877
H_9	197	353	1.792	0.920
H_{10}	190	370	1.947	1
H_{11}	270	350	1.296	0.666
H_{12}	250	362	1.448	0.744

To draw such comparison, we implicitly accept that fewer staff and more admitted patients are desirable. In general, there are two mutually exclusive items for measuring the efficiency score: *input* and *output.* The smaller input and larger output amounts are preferable. According to this definition, in this numerical example, the number of staff and the number of patients are input and output, respectively.

In the aforementioned discussion, we have just relatively compared the performance of two dominating and dominated hospitals and could not measure their efficiency scores. In addition, such analysis fails to compare two arbitrary hospitals. To measure the efficiency score of all hospitals, the following ratio can be utilized:

$$\text{Efficiency} = \frac{\text{Output}}{\text{Input}}$$

This ratio can be interpreted as a patient per staff (*output per input*) and it can be found that more output amounts and/or less input amounts lead to more efficiency score.

In Table 1–1, the fourth column indicates the ratio of patients to staff. The normalized efficiency score are presented in the last column, that is, the maximum one, H_{10}, becomes unity. The main difference between these measures will be discussed subsequently. Based on this ratio definition for measuring efficiency, as we expect, the efficiency score of H_5, 0.736, is larger than H_1, 0.659 and also H_8 with $e_8 = 0.877$ is more efficient than H_{12} with $e_{12} = 0.744$

Note that the efficiency measured in this approach is not absolute and hence if a hospital is removed or added, then the efficiency *can* be changed. It is easy to show that the efficiency scores will not change if an inefficient hospital is removed or a hospital with $y \leq 1.947x$ is added. On the other hand, with removing H_{10} or adding a hospital with $y > 1.947x$, the efficiency scores will change.

To generalize the aforementioned method, suppose that there are *n* homogeneous organizations for evaluating in which each organization utilizes a single input to produce a single output. We call these similar organizations as decision making units (DMUs[1]) and denote by DMU_j $(j = 1, \dots, n)$. With these assumptions the efficiency score of the under evaluation unit, $\mathrm{DMU}_{o \in \{1,\dots,n\}}$, can be measured as

$$e_o = \frac{\frac{y_o}{x_o}}{\max\left\{\frac{y_j}{x_j} : j = 1, \dots, n\right\}} \tag{1.1}$$

where $x_j > 0$ and $y_j > 0$ are input and output of $\mathrm{DMU_j}$, respectively. Clearly, efficiency score, e_j, is a positive number which is less than or equal to 1, i.e. $\forall j\ e_j \in (0, 1]$ and hence DMU_k is *efficient* if and only if $e_k = 1$; otherwise it is *inefficient*. In this method, since the performance of a DMU is relatively measured, e_j can be called *relative efficiency* score of DMU_j and there is always at least one efficient unit. Consider the following set

$$EF = \left\{(x, y) : \frac{y}{x} = \underset{j=1,\dots,n}{\text{maximum}} \left\{\frac{y_j}{x_j}\right\}\right\}$$

Let $\mathrm{DMU}_k = (x_k, y_k) \in EF$. If $e_k = 1$ then EF is called *efficient frontier* since it involves all the efficient points. We will discuss more about the efficient frontier shortly.

Note that the above formula (1.1) is a *ratio of ratios form* and hence is able to present the two properties: *units invariant* and *constant returns to scale* (CRS). A measure is called units invariant if it is independent of the units of measurements. In other words, the efficiency scores remain unchanged if one multiplies *all* inputs by a positive constant α and *all* outputs by a positive constant β:

$$\frac{\frac{y_o}{x_o}}{\max\left\{\frac{y_j}{x_j} : j = 1, \dots, n\right\}} = \frac{\frac{\alpha y_o}{\beta x_o}}{\max\left\{\frac{\alpha y_j}{\beta x_j} : i = 1, \dots, n\right\}}$$

If inputs and outputs of a DMU are multiplied by a positive constant α and the efficiency scores remain unchanged, the unit operates under CRS assumption.

---

[1] In general, a DMU converts input(s) into output(s) and whose efficiency score is desirable. In managerial applications, DMUs may include hospitals, banks, schools and et cetera. In engineering, DMUs may take such forms as airplanes or their components such as jet engines.

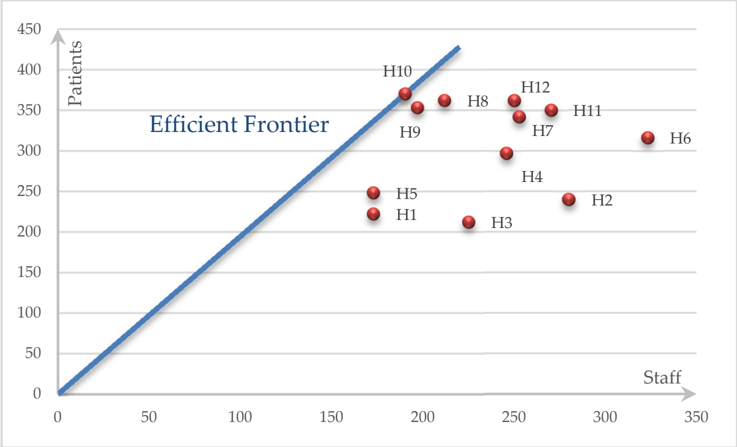


Figure 1–1 Data set and efficient frontier

If we return to the numerical example we can notice that H_{10} is efficient and relative to this hospital, the worst hospital (H_3) attains only 0.440 of efficiency score. In a similar manner, the efficiency score of other hospitals can be interpreted. In addition, if the number of patients were restated in units of persons/month (instead of 100 persons/month), then the ratio of output to input, the fourth column in Table 1–1, would multiply by 100 whereas the ratio of ratios, the last column, remains unchanged. Furthermore, multiplying the number of staff and the number of patients of a DMU by a positive number does not change the obtained efficiency score.

To provide some geometric insights, we plot number of staff and number of patients on the horizontal and vertical axes, respectively. The line that connects each observation to the origin shows the set of points with an identical slope (i.e., ratio of output to input). Considering the CRS assumption, the efficiency score of all points on the line is the slope of the line, which is a constant measure.

The efficient frontier is the highest slope ($y = 1.947x$) and as a result this frontier *envelops* all *data*[2]. The following figure depicts the data set and the efficient frontier.

Mathematically, DMU_p dominates DMU_q when $x_p \leq x_q$ and $y_p \geq y_q$ and if we also have $x_p < x_q$ or $y_p > y_q$, then DMU_p *strictly* dominates DMU_q. A DMU is efficient if other DMUs cannot strictly dominate it. To improve the efficiency of an inefficient unit such as H_1, we look for a point on the efficient frontier that strictly dominates H_1. Figure 1–1 depicts the set of all points on the efficient frontier that dominate H_1, i.e. the line segment AB. This line is called *project points* because one of them

[2] The name Data Envelopment Analysis (DEA) stems from this property.

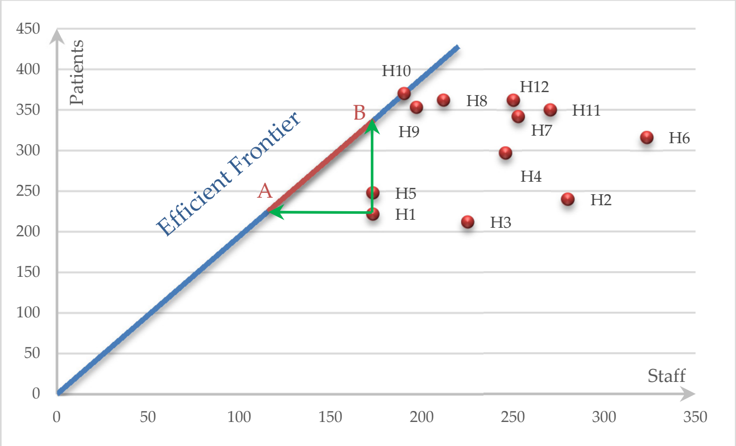


Figure 1–2 Projection Points

must be selected as a target for improving H1. Note that it is assumed that increasing in input and/or decreasing in output are not preferable and hence it is not allowed. Considering the line $y = 1.947x$ as the efficient frontier, the efficient point A with coordinates $(\frac{222}{1.947}, 222)$ is achieved by decreasing the input of H1 (*input orientation*) and similarly the efficient point B with coordinates $(173, 173 \times 1.947)$ is gained by increasing the output of H1 (*output orientation*).

As it can be extracted from Figure 1–2, the set of all points that dominates H1 is the triangle ABH1 and subsequently the efficiency score of H1 is equal to the ratio of $x_A = 114.02$ to $x_{H_1} = 173$ or equivalently the ratio of $y_{H_1} = 222$ to $y_B = 336.83$:

$$e_{H_1} = \frac{114.02}{173} = \frac{222}{336.83} = 0.659$$

The DM believes that obtaining more details about the data set of hospitals leads to more accurate efficiency scores. In such case, we can consider three following scenarios:

1. Decomposing the number of staff into the number of doctors and the number of nurses, which implies two inputs and one output case.
2. Decomposing the number of patients into the number of inpatients and the number of outpatients that gives us one input and two outputs case.
3. Decomposing the number of staff and the number of patients, simultaneously, (scenario 1 plus scenario 2) which leads to two inputs and two outputs case.

To gain basic insights, we discuss these scenarios in following sections:

1. 2 Two inputs and one output

Suppose that the number of staff is decomposed into the number of doctors and the number of nurses for all hospitals. The columns in Table 1–2 labeled *Doctors* and *Nurses* show this decomposition, for example in H_1, the number of staff, 173, is equal to the number of doctors plus the number of nurses, 25 + 148. We apply the CRS property and normalize the input values to get 1 unit for output which is shown in the columns labeled *Normalized Inputs* and *Unitized Output* in Table 1–2. The efficiency scores are shown in the last column and the calculation method will be discussed successively.

With this slight manipulation, we can plot the new data set on new inputs axes. Since the new output is unitized to 1, we can ignore it and just consider *unitized axes*, Doctors/Patients (x) and Nurses/Patients (y), in Figure 1–3. The figure illustrates the intersection of efficient frontier in R^3 when Patients=1. The line connecting H_8, H_9 and H_{10} indicates the non-dominated space to be built the efficient frontier. Notice that the vertical and horizontal dashed lines and the efficient frontier envelop all data set. H_8, H_9 and H_{10} are the efficient hospitals, which illustrate that more efficient hospitals *might* be obtained if we use more input and output factors.

There are many alternative methods to project an inefficient hospital on the efficient frontier. One method is to radially decrease the input values, as much as possible, meaning that it identifies a set of points with an identical ratio of inputs[3]. Having the same ratio of inputs for a projection point implies that all inputs can be simultaneously reduced without altering the mix (=proportions) in which they are utilized. Therefore, to improve an inefficient hospital, H_5, we connect it to the origin by a green solid line. This line shows the set of all points with the same mix of H_5's inputs and hence an efficient point in this line indicates a suitable projection point for H_5, as specified by P in Figure 1–4.

The efficiency score of H_5 can be measured by comparing the current position, H_5, with the projection point P, as below

$$e_{H_5} = \frac{d(O,P)}{d(O,H_{5)}} = \frac{\sqrt{0.0713^2 + 0.465^2}}{\sqrt{0.093^2 + 0.605^2}} = 0.769$$

where $d(O,P)$ and $d = (O,H)$ are distance from origin to P and H_5, respectively. The coordinates of P can be easily calculated from intersection of $y = \frac{0.605}{0.093}x$, associated with the segment line OH_5, and $y = -2.834x + 0.667$ associated with the segment line H_9H_{10}. Since the projection point P is on the line connecting H_9 and H_{10}, the set containing these efficient hospitals is called *reference set* for H_5. In

[3] In this example, the improvement of an inefficient hospital cannot be implemented by increasing the outputs because it is assumed that they are be fixed at 1.

Table 1–2 Two inputs and one output case

Hospitals	Inputs		Output	Normalized Inputs		Unitized Output	Efficiency
	Doctors	Nurses	Patients	Doctors/Patients	Nurses/Patients	Patients/Patients	
H1	25	148	222	0.113	0.667	1	0.677
H2	46	234	240	0.192	0.975	1	0.441
H3	32	193	212	0.151	0.910	1	0.499
H4	39	207	297	0.131	0.697	1	0.624
H5	23	150	248	0.093	0.605	1	0.769
H6	36	287	316	0.114	0.908	1	0.542
H7	50	203	342	0.146	0.594	1	0.724
H8	16	196	323	0.050	0.607	1	1
H9	21	176	353	0.059	0.499	1	1
H10	31	159	370	0.084	0.430	1	1
H11	45	225	350	0.129	0.643	1	0.668
H12	43	207	362	0.119	0.572	1	0.752

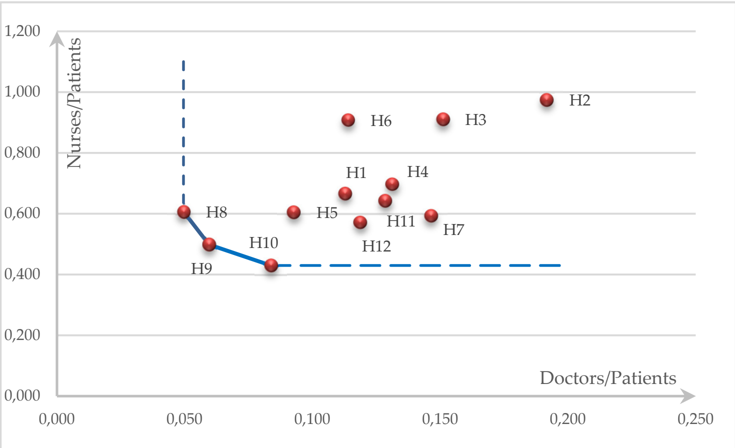


Figure 1–3 Two inputs and one output case

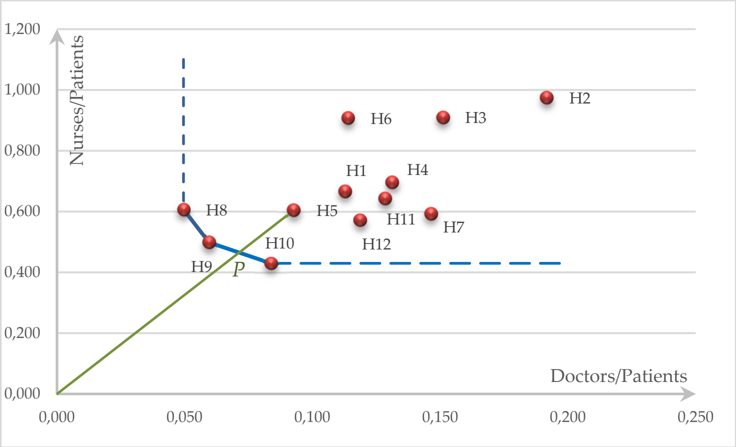


Figure 1–4 Improvement of H_5

The efficiency score of H_5 is equal to 0.769, meaning that reducing the coordinates, or equivalently both inputs of this hospital by 0.769 leads to a point on the efficient frontier. In other words, 0.769(0.093, 0.605) is equal to the coordinates of P, which is located on the efficient frontier.

In Figure 1–4, it is impossible to figure out the output orientation improvement because the number of patients is fixed at one. One approach to show the output orientation is to unitize one input and plot the new normalized data. Toward this end, we consider one input such as Nurse that are one for all hospitals as well as normalizing Doctors and Patients. Although the efficiency scores will remain unchanged the form of efficient frontier and subsequently the way of efficiency measurement will be changed. Table 1–3 demonstrates the alternative normalized data set.

Figure 1–5 shows the new normalized data set where the new horizontal and vertical unitized axes are Doctors/Nurses (x) and Patients/Nurses (y), respectively. The efficient frontier in this figure is the intersection of efficient frontier in R^3 when Nurses=1. As a result, the dashed segment line OH_8 and the vertical segment line AH_5 in Figure 1–5 correspond to vertical dashed line and the segment line OH_5 in Figure 1–3, respectively.

In this figure, the vertical segment line AH_5 identifies the set of all points with the same ratio of H_5's inputs, 1.653. As we expect, the projection point of H_5 is located on the line connecting H_9 and H_{10}. The optimal achievement of H_5's output is as follows:

$$\frac{y_p}{y_{H_s}} = \frac{2.1502}{1.653} = 1.3$$

Table 1–3 Data with unitized Nurses

Hospitals	Normalized Input Doctors / Nurses	Unitized Input Nurses / Nurses	Normalized Output Patients / Nurses	Efficiency
H_1	0.169	1	1.500	0.677
H_2	0.197	1	1.026	0.441
H_3	0.166	1	1.098	0.499
H_4	0.188	1	1.435	0.624
H_5	0.153	1	1.653	0.769
H_6	0.125	1	1.101	0.542
H_7	0.246	1	1.685	0.724
H_8	0.082	1	1.648	1
H_9	0.119	1	2.006	1
H_{10}	0.195	1	2.327	1
H_{11}	0.200	1	1.556	0.668
H_{12}	0.208	1	1.749	0.752

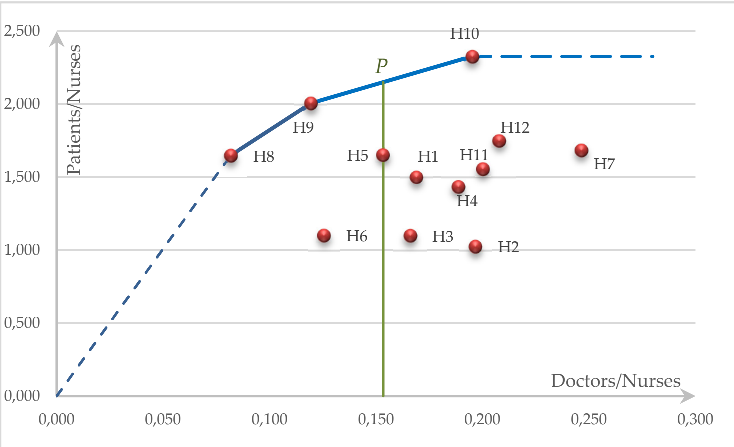


Figure 1–5 Alternative perspective of two inputs and one output case

where $p = (0.153, 2.1502)$ is obtained from the intersection of $y = 4.248x + 1.299$, the segment line H_9H_{10}, and $x = 0.153$. Put differently, if H_5 increases its output by 1.3, then it would be efficient. It should be noticed that the efficiency score of H_5 is equal to $\frac{1}{1.3} = 0.769$ which practically shows the CRS property.

As a matter of fact, vertically moving up from H_5 to the efficient frontier in Figure 1–4 (Doctors/Nurses, Patients/Nurses) space is equivalent to radially moving from H_5 to the efficient frontier in Figure 1–5 (Doctors/Patients, Nurses/Patients) space and these movements are two specific projections from a general movement in (Doctors, Nurses, Patients) space.

1. 3 One input and two outputs

Now let us consider the second scenario and suppose that the number of patients is decomposed into the number of inpatients and the number of outpatients. The third and fourth columns of Table 1–4 show the decomposition, for example in H_1, the number of patients, 222, is equal to the number of inpatients plus the number of outpatients, 33 + 189. The unitized input and normalized outputs are shown in the 5th–7th columns of this table. The efficiency scores are reported in the last column and the utilized geometrical method will be discussed successively.

Figure 1–6 illustrates one input and two outputs case geometrically where the horizontal and vertical unitized axes are inpatients/staff and outpatients/staff, respectively. There are three efficient hospitals, i.e. H_9, H_{10} and H_9, and the lines that connect these hospitals are efficient frontier.

To improve an inefficient hospital we radially increase its outputs (remember that inputs are fixed). Consider an inefficient hospital, H_7, and the segment line that connects origin to H_7 and crosses the efficient frontier at A. This segment line, OA, which shows the set of all points with the same ratio of H_7's outputs, is plotted in Figure 1–7.

Table 1–4 One input and two outputs case

Hospitals	Input	Outputs		Unitized input	Normalized output		Efficiency
	Staff	Inpatients	Outpatients	Staff/Staff	Inpatients/Staff	Outpatients/Staff	
H_1	173	33	189	1	0.191	1.092	0.687
H_2	280	65	175	1	0.232	0.625	0.513
H_3	225	70	142	1	0.311	0.631	0.633
H_4	246	63	234	1	0.256	0.951	0.634
H_5	173	87	161	1	0.503	0.931	1
H_6	323	91	225	1	0.282	0.697	0.606
H_7	253	84	258	1	0.332	1.020	0.766
H_8	212	68	294	1	0.321	1.387	0.887
H_9	197	33	320	1	0.168	1.624	1
H_{10}	190	75	295	1	0.395	1.553	1
H_{11}	270	73	277	1	0.270	1.026	0.675
H_{12}	250	69	293	1	0.119	0.572	0.751

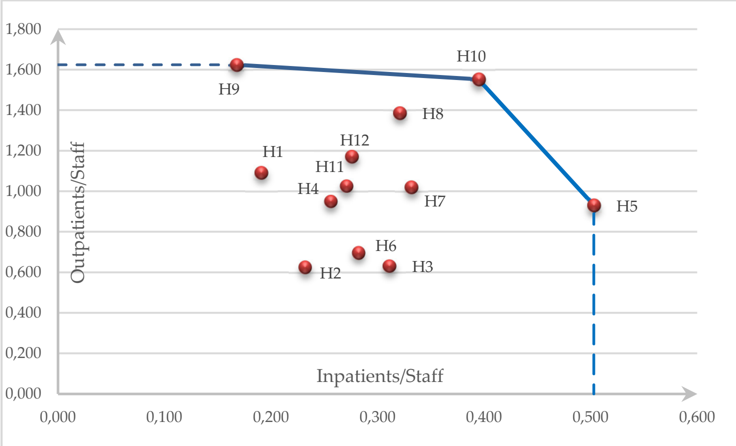


Figure 1–6 One input and two outputs case

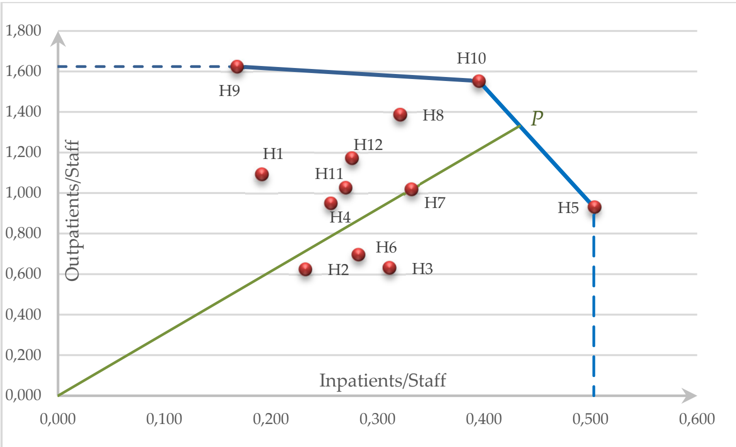


Figure 1–7 Improvement of H7

Similarly, the efficiency score of H_7 can be measured using the formula below

$$e_{H_7} = \frac{d(O, H_7)}{d(O, P)} = \frac{\sqrt{0.332^2 + 1.020^2}}{\sqrt{0.433^2 + 1.331^2}} = 0.766$$

The coordinates of P can be calculated from intersection of $y = \frac{1.020}{0.332}x$ and $y = -5.751x + 3.828$. The reference set for the inefficient hospital H_7 involves H_5 and H_{10}.

To implement the input orientation improvement under CRS, we normalize the input value staff and the output value inpatients to get the unity value for the output outpatients, which is shown in Figure 1–8.

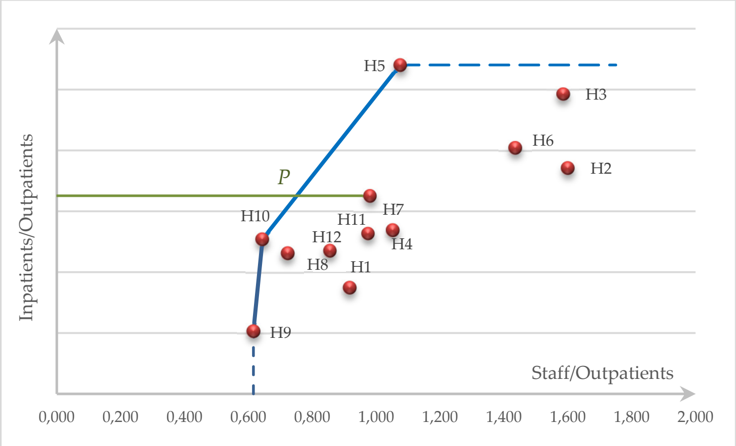


Figure 1–8 Alternative perspective of one input and two outputs case

Horizontal moving from H7 to the efficient frontier in Figure 1–8 is equivalent to radially moving from this hospital to the efficient frontier in Figure 1–7. The similar approach can be utilized to perform all necessary computations and obtain instructive details.

1. 4 Two inputs and two outputs

Now, suppose that for each DMU there are two inputs, the number of doctors and nurses, and two outputs, the number of inpatients and outpatients as demonstrated in Table 1–5. One approach to measure the efficiency scores in this case, might unitize one input or one output. We then plot the data set by taking three normalized items as axes. However, it is not easy to analyze a three-dimensional graph and more importantly this approach cannot be extended if more than four items exist or DMUs operate under *variable returns to scale* (VRS) assumption. One practical method to tackle this issue is to assign a *fixed* weight to inputs and outputs in order to obtain one input as the weighted sum of inputs, and one output as the weighted sum of outputs. Therefore, the one-input-one-output method can be applied as discussed above. Nevertheless, the main issue in this approach is how to select weights, because various weights might lead to different efficiency scores. To illustrate the role of selected weights, we consider the numerical example with 12 hospitals in terms of two decomposed inputs and two decomposed outputs as exhibited in Table 1–5.

The imposition of the fixed weights is very critical since the efficiency score is driven from these weights. Having the fixed weights, the ratio of weighted sum of outputs to weighted sum of inputs for all hospitals can be calculated and then the normalized ratio would then provide an index for evaluating efficiency scores. Let v_1 and v_2 be the weights for Doctors and Nurses, respectively. The data set in Table

Table 1–5 Two inputs and two outputs case

Hospital	Inputs		Outputs	
	Doctors	Nurses	Inpatients	Outpatients
H_1	25	148	33	189
H_2	46	234	65	175
H_3	32	193	70	142
H_4	39	207	63	234
H_5	23	150	87	161
H_6	36	287	91	225
H_7	50	203	84	258
H_8	16	196	68	294
H_9	21	176	33	320
H_{10}	31	159	75	295
H_{11}	45	225	73	277
H_{12}	43	207	69	293

1–4 will be obtained if we select $(v_1, v_2) = (1,1)$. Similarly, selecting $(u_1, u_2) = (1,1)$ leads to the data set in Table 1–2 where u_1 and u_2 are the weights for inpatients and outpatients, respectively, also the unit weights $(v_1, v_2, u_1, u_2) = (1,1,1,1)$ result in the data in Table 1–1.

Table 1–6 reports the different efficiency scores when various fixed weights are selected. The efficiency scores in the second column are obtained by fixing $(v_1, v_2, u_1, u_2) = (5,1,1,4)$ which means that a doctor is weighted five times more than a nurse and also an outpatient is weighted four times more than an inpatient. The efficiency scores presented in the third and fourth column are the fixed weights (2.5,0.3,1.5,4.5) and (0.3,1,0.3,1), respectively. As can be seen, different fixed weights lead to different efficiency scores. A main question here is which weights should be selected? Which one is better? Obviously, the best (optimal) weights must be selected not the better one; otherwise, it is not clear how much the efficiency scores are due to assigning a non-optimal weights and how much inefficiency is associated with the data. To have the optimal weights we must have some criteria which lead to different types of optimization problems: A multi-objective mathematical programming approach can be developed if maximizing the individual efficiency score of all units is desirable. An integrated minimax mathematical programing method can be formulated if minimizing the deviation from efficiency score of the worst inefficient DMU is preferred. Another criterion might be maximizing the sum of efficiency score of all DMUs. The last column in Table 1–6, labeled *CSW*, indicates the efficiency score via an aggregation minimax mathematical approach.

Table 1–6 Various weights

Hospitals	(5,1,1,4)	(2.5,0.3,1.5,4.5)	(0.3,1,0.3,1)	CSW
H_1	0.619	0.584	0.678	0.662
H_2	0.353	0.331	0.416	0.440
H_3	0.387	0.374	0.426	0.486
H_4	0.532	0.498	0.613	0.621
H_5	0.590	0.578	0.632	0.741
H_6	0.454	0.452	0.449	0.509
H_7	0.527	0.480	0.689	0.689
H_8	0.965	1	0.830	0.895
H_9	1	0.981	0.959	0.932
H_{10}	0.855	0.797	1	1
H_{11}	0.562	0.522	0.664	0.665
H_{12}	0.629	0.581	0.756	0.742
sum	7.473	7.180	8.113	8.382

The values at the bottom of each column show the sum of various efficiency scores via different weights. The aggregation minimax method gives a set of efficiency scores with higher value of the sum of efficiency than the other methods. Most importantly, as will be discussed in the next section, an advantage of the optimization approaches is that it is not a need to pre-select the fixed weights for inputs and outputs.

The weights obtained from all the aforementioned optimization methods are common to all DMUs and hence are called *common set of weights* (CSW). Beside these approaches, there are some others that allow the weights to vary from one DMU to another. The maximum efficiency score for each DMU obtains with such flexible weights. More details about the approaches and the proposed models are explained in the succeeding chapters. However, to illustrate the benefit of flexible weights, Table 1–7 exhibits optimal weights and the efficiency scores that are calculated by the basic DEA model, i.e. CCR[4].

In this table, the last column which is labeled *CCR* shows the efficiency score and v_1^*, v_2^*, u_1^* and u_2^* are the optimal weights for the number of doctors, nurses, inpatients and outpatients, respectively. There are four efficient DMUs, i.e. H_5, H_8, H_9 and H_{10}, in this method while in the previous approaches only one efficient DMU was determined. In some cases, some optimal weights are zero which means the corresponding item must be ignored to gain the maximum efficiency score. For

[4] Originated by Charnes, Cooper and Rhodes in 1978.

Table 1–7 The CCR efficiency scores and optimal weights

Hospitals	v_1^*	v_2^*	u_1^*	u_2^*	**CCR**
H_1	1.80×10^{-03}	6.45×10^{-03}	0	3.67×10^{-03}	0.693
H_2	0	4.27×10^{-03}	5.86×10^{-03}	8.12×10^{-03}	0.523
H_3	0	5.18×10^{-03}	7.11×10^{-03}	9.85×10^{-03}	0.638
H_4	9.91×10^{-03}	2.96×10^{-03}	5.38×10^{-03}	1.27×10^{-03}	0.636
H_5	1.47×10^{-02}	4.41×10^{-03}	7.99×10^{-03}	1.89×10^{-03}	1
H_6	2.31×10^{-02}	5.81×10^{-04}	7.12×10^{-03}	0	0.648
H_7	0	4.93×10^{-03}	6.76×10^{-03}	9.36×10^{-04}	0.809
H_8	2.87×10^{-02}	2.76×10^{-03}	2.76×10^{-03}	2.76×10^{-03}	1
H_9	1.20×10^{-02}	4.25×10^{-03}	2.82×10^{-03}	2.83×10^{-03}	1
H_{10}	2.70×10^{-03}	5.76×10^{-03}	2.70×10^{-03}	2.70×10^{-03}	1
H_{11}	0	4.44×10^{-03}	6.10×10^{-03}	8.45×10^{-04}	0.679
H_{12}	0	4.83×10^{-03}	0	2.60×10^{-03}	0.763

instance, the optimal weight for the first output of H_1 is zero ($u_1^* = 0$) and hence the efficiency score for H_1 (0.693) is calculated in the ignorance of the number of inpatients item for all hospitals.