DATA SCIENCE

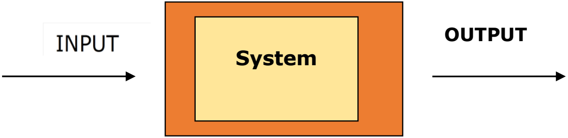
Comparative evaluation of firm performance by Data Envelopment Analysis

Antonio VIOLI

antonio.violi@unirc.it

Introduction

- In different application contexts, often arises the need to assess and compare the efficiency of complex organizations, such as companies, university departments, hospitals, bank branches.
- Intuitively the efficiency is related to the ability to well manage the available resources
- In the following, we shall represent a system as a black box that receives some inputs to produce some outputs



Introduction

- Commonly, the efficiency is measured by taking into account the relationship that exists between the total quantity of products/services realized (output) and the amount of resources used (input).
- In quantitative terms

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

- Efficiency can be improved by increasing the output level for a fixed input level, or viceversa, by reducing the input level for a given output level

Different kinds of efficiency

We may distinguish different efficiency measures:

- **Technical Efficiency**
  It measures the performance of a system considered as a "physical" conversion process that transforms the inputs (resources: human, technological, material, financial) to output (products or services) made.
  The technical efficiency is therefore understood as "efficiency" of the production process.

Different kinds of efficiency

- **Allocative Efficiency**
  Assuming that the production process is technically efficient, the allocative efficiency refers to the minimization of costs through the proper choice of the factors used as input to obtain a given level of output.
  The allocative efficiency is related to the management policy that maximizes the output level by allocating the inputs in the best possible way
- **Economic Efficiency**
  It refers jointly the technical and allocative efficiency.

Technical efficiency

- We shall focus on technical efficiency and assume that a complex system can be seen as a set of parts, called Decision Making Unit (DMU).
- To ensure that the evaluation process can be considered reliable, we require that the units to be evaluated are:
  - **Homogeneous**
    the DMUs must use the same type of resources to produce the same output and must be similar (size, sales volume)
  - **Independent**
    the behavior of a DMU should not influence that of the others
  - **Autonomous**
    each unit must have the ability to decide for itself how to use its resources.

Illustrative example

Let us suppose to consider 8 DMUs each of which characterized by a single input and a single output. Suppose that the units are representative of 8 sale points of the same store chain and to consider as input the number of employees and as output the sale volume.

Branch	A	B	C	D	E	F	G	H
Employees	2	3	3	4	5	5	6	8
Sale volume	1	3	2	3	4	2	3	5

The owner wants to know which store is efficient and which store is inefficient. He is also interested to benchmark the best store so that he can suggest improvements for the inefficient stores by comparing inefficient stores with the efficient ones.

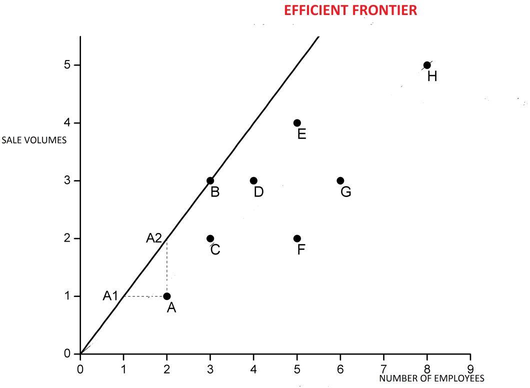
Illustrative example

By computing the efficiency as the ratio output on input, we find that the most efficient store is B, whereas the less efficient is F.

Branch	A	B	C	D	E	F	G	H
Employees	2	3	3	4	5	5	6	8
Sale volume	1	3	2	3	4	2	3	5
Sale/Employee	0.5	1	0.66	0.75	0.8	0.4	0.5	0.625

Illustrative example

- We may plot the input and output of each store on a graph
- If we extend the line segment from origin to store B, it will envelope all the other points because it hasthe maximum slope and efficiency. This line segment is called **efficient frontier**



The efficient frontier

- The Efficient Frontier envelopes all the other data points
- The technique that we are going to introduce is called **Data Envelopment Analysis** (DEA, for short).
  - The frontier line displays the performance of the best store in the comparison to the others.
  - The efficiency of the other stores can be measured by considering the deviation of the points from the frontier line.
  - Efficient frontier serves as Benchmark.

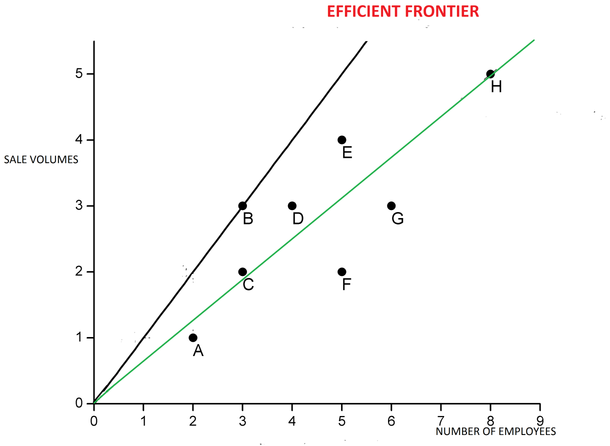
The efficient frontier

- Analyzing the figure you can get useful information on how you can make efficient an inefficient unit
- Consider for example the unit A. Its efficiency can be improved in several ways
- We may reduce the input keeping fixed the output level (point A_1)
- We may increase the output by keeping fixed the input level (point A_2)
- In general, all the points on the segment between A_1 and A_2 have the opportunity to make efficient the unit A.

Remarks

- Using the data shown in the figure, you might think to perform a linear regression, that is look for the line that best interpolates all the points
- This type of analysis, typically used in statistics, provides an indication of the average behavior, that could be used to classify, for example, as efficient the points above the straight line and as inefficient those points below the line
- The difference between the DEA and the linear regression is evident
- DEA identifies the most efficient unit, therefore, the best performance, whereas the regression technique provides a sort of average measure

Comparison with the regression analysis

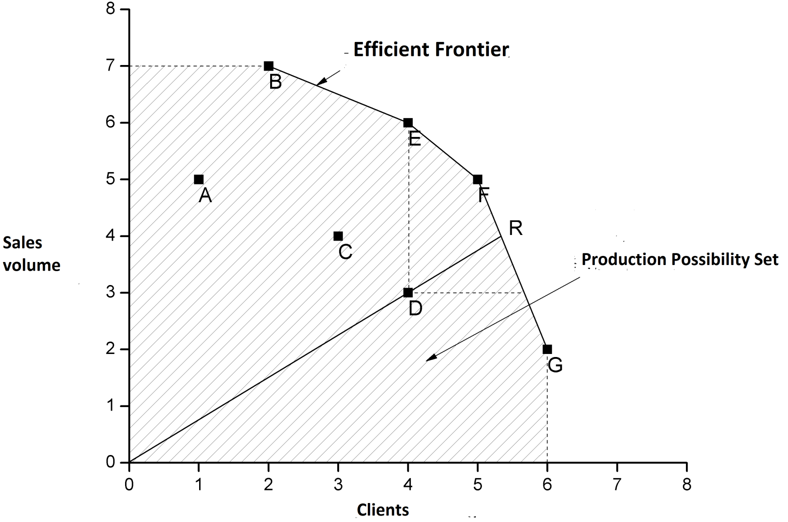


Illustrative example: multiple output one input

- A retail giant has 7 stores in a metro city. The owner of the retail giant wants to compare the performance of the retail stores. The owner would like to consider as input the number of employees (equal for all the stores) and as outputs the number of clients visiting the stores and the sale volumes.

Stores	A	B	C	D	E	F	G
clients	1	2	3	4	4	5	6
Sale volume	5	7	4	3	6	5	2

Illustrative example: multiple output one input



Illustrative example: multiple output one input

- Similar to the previous case, the efficiency of a particular store can be obtained by connecting a line from the origin to the point of the store, crossing the efficiency frontier.
- Efficiency of D = distance (O,D) / distance (O,R) = $5/(20/3) = 0.75$.
- If we consider the reciprocal of the previous value (1.33), we have an indication of how we should increase both outputs of unit D to make it efficient.
- Multiplying this ratio by the coordinates of the point D, we obtain the coordinates of the point R on the efficient frontier.
- We observe that the units F and G represent the reference units for D (P is located on the segment FG).
- There are, however, other possibilities to improve the efficiency D, such as increasing an output keeping fixed the other settings.

The Data Envelopment Analysis Technique

- When multiple inputs and multiple outputs are considered, the efficiency can not be determined as in the previous examples
- We recall that efficiency is defined as the ratio output on input.
- In the general case, we have a weighted sum and, thus, we need to determine the weights associated with the inputs and the outputs
- The weights associated with input can be considered as a cost that should be reduced, whereas those associated with the outputs represent a value that should be increased
- The optimal weights and, thus, the efficient frontier can be determined by applying the DEA technique which is based on the definition and solution of LP models

The Data Envelopment Analysis Technique

- We assume to consider a complex organization that consists of K DMU
- We assume that each unit k can be described in terms of m inputs (consumed resources) and n outputs (realized products or services)
- We denote by
  - x_i^k, with $i = 1, \ldots, m$, the level of input i used by k
  - y_j^k, with $j = 1, \ldots, n$, the level of output j produced by k
- In the following we shall assume that all the parameters are non negative (if this is not the case some specialized technique can be used)

The DEA Technique

We denote by

- w_i^k the weight associated with input i, $i = 1, \ldots, m$
- t_j^k the weight associated with output j, $j = 1, \ldots, n$

The efficiency of the DMU k is defined as

$$E^k = \frac{\sum_{j=1}^{n} y_j^k t_j^k}{\sum_{i=1}^{m} x_i^k w_i^k}$$

The efficiency is related to the capacity to better manage the available resources in such a way to realize a greater level of outputs by using the input level, or to use a lower level of inputs to produce greater amounts of outputs
Each DMU will have to determine the best value of the weights in comparison with the other DMUs in order to achieve its maximum performance.

The DEA Technique

For each DMU k, we define the model

$$\max \quad E^k = \frac{\sum_{j=1}^{n} y_j^k t_j^k}{\sum_{i=1}^{m} x_i^k w_i^k}$$

$$\frac{\sum_{j=1}^{n} y_j^l t_j^k}{\sum_{i=1}^{m} x_i^l w_i^k} \leq 1 \quad l = 1, \ldots, K$$

$$t_j^k \geq 0 \quad j = 1, \ldots, n$$

$$w_i^k \geq 0 \quad i = 1, \ldots, m$$

Let E^{k*} the optimal objective function of the previous problem

$$E^{k*} \begin{cases} =1 & \text{DMU k is efficient} \\ <1 & \text{DMU is inefficient} \end{cases}$$

In this second case, there exists other DMUs which are efficient, even though the model has determined the weights that optimize the efficiency of DMU k