DATA SCIENCE

Comparative evaluation of firm performance by Data Envelopment Analysis

Antonio VIOLI

antonio.violi@unirc.it

The DEA Technique

For each DMU k, we define the model

$$
\begin{aligned}
\max \quad & E^k = \frac{\sum_{j=1}^{n} y_j^k t_j^k}{\sum_{i=1}^{m} x_i^k w_i^k} \\
& \frac{\sum_{j=1}^{n} y_j^l t_j^k}{\sum_{i=1}^{m} x_i^l w_i^k} \leq 1 \quad l = 1, \ldots, K \\
& t_j^k \geq 0 \quad j = 1, \ldots, n \\
& w_i^k \geq 0 \quad i = 1, \ldots, m
\end{aligned}
$$

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

Remarks

- **The weight issue**
  As we have seen, the weights are defined as non negative variables. Typically a given threshold δ, greater than 0, but infinitesimal, is considered. This assures that all the inputs and outputs are taken into account in the evaluation process.
- **The nature of the model**
  The model introduced above presents a fraction and, as such, is nonlinear. In the literature, different techniques have been proposed with the aim of deriving a linear model and different versions of the basic model have been studied

The CCR version: Input Oriented

This version is due to Charnes, Cooper and Rhodes (CCR).
For the sake of simplicity, we shall omit the superscript k in the weights

$$\max z = \sum_{j=1}^{n} y_j^k t_j$$

$$\sum_{i=1}^{m} x_i^k w_i = 1$$

$$\sum_{j=1}^{n} y_j^l t_j - \sum_{i=1}^{m} x_i^l w_i \leq 0 \quad l = 1, \ldots, K$$

$$w_i \geq \delta \quad i = 1, \ldots, m$$

$$t_j \geq \delta \quad j = 1, \ldots, n$$

The CCR version: Input Oriented

- The optimal value of the objective function corresponds to the efficiency of DMU k
- If $z^* = 1$ then DMU k is efficient
- For an inefficient DMU, we may determine the Peer Group o Reference Set for k

$$RF(k) = \left\{ l \;\middle|\; \sum_{j=1}^{n} y_j^l t_j^* - \sum_{i=1}^{m} x_i^l w_i^* = 0 \right\}$$

The CCR version: Output Oriented

$$
\begin{aligned}
\min z = \quad & \sum_{i=1}^{m} x_i^k w_i \\
& \sum_{j=1}^{n} y_j^k t_j = 1 \\
& \sum_{j=1}^{n} y_j^l t_j - \sum_{i=1}^{m} x_i^l w_i \leq 0 \quad l = 1, \ldots, K \\
& w_i \geq \delta \quad i = 1, \ldots, m \\
& t_j \geq \delta \quad j = 1, \ldots, n
\end{aligned}
$$

The efficiency level is $1/z^*$

Example

Super labs are in Research and Development of electronic goods. The founders, in the year 2001 decided to set-up their own company to produce consumer electronic goods that use the state-of-the art technologies that have been developed by them. The company has set up business in North America and quickly moved to Asia, Europe, Africa and South America. The main products of the company are its impressive range of Smart phones, Tablet PCs and Laptops. At the beginning of the year 2011, the CEO- global operations of the company quickly pulled reports of Inputs and Outputs that are being used and produced at various regions in which company operates.

Example

	INPUTS($ in millions)		OUTPUTS (numbers in '000)		
DMU	Product Development	Marketing	Smart Phones	Tablets	Laptops
N. America	5	14	9	4	16
Europe	10	18	8	2	9
Asia	9	16	9	4	10
Africa	7	12	6	1	8
S. America	9	15	10	4	14

Example

The data tells him how much each of the regions consume for the two main activities, namely Product development and Marketing, and how many units of its products are being sold. However, he is unable to figure out, which are the best performing units and how much of increase or decrease the Regional directors are to be advised to make on investments. One of his analysts recommends the use of Data Envelopment Analysis to analyze the performance of the different units and then make a conclusion.

Example

For each DMU a different DEA model is solved.
For N. America, the model is the following:

$$
\begin{aligned}
\max z \quad & = 9t_1 + 4t_2 + 16t_3 \\
& 5w_1 + 14w_2 = 1 \\
& 9t_1 + 4t_2 + 16t_3 - 5w_1 - 14w_2 \leq 0 \\
& 8t_1 + 2t_2 + 9t_3 - 10w_1 - 18w_2 \leq 0 \\
& 9t_1 + 4t_2 + 10t_3 - 9w_1 - 16w_2 \leq 0 \\
& 6t_1 + 1t_2 + 8t_3 - 7w_1 - 12w_2 \leq 0 \\
& 10t_1 + 4t_2 + 14t_3 - 9w_1 - 15w_2 \leq 0 \\
& w_i \geq \delta \quad i = 1, 2 \\
& t_j \geq \delta \quad j = 1, 2, 3
\end{aligned}
$$

Example

For Europe, the model is the following:

$$
\begin{aligned}
\max z \quad & = 8t_1 + 2t_2 + 9t_3 \\
& 10w_1 + 18w_2 = 1 \\
& 9t_1 + 4t_2 + 16t_3 - 5w_1 - 14w_2 \leq 0 \\
& 8t_1 + 2t_2 + 9t_3 - 10w_1 - 18w_2 \leq 0 \\
& 9t_1 + 4t_2 + 10t_3 - 9w_1 - 16w_2 \leq 0 \\
& 6t_1 + 1t_2 + 8t_3 - 7w_1 - 12w_2 \leq 0 \\
& 10t_1 + 4t_2 + 14t_3 - 9w_1 - 15w_2 \leq 0 \\
& w_i \geq \delta \quad i = 1, 2 \\
& t_j \geq \delta \quad j = 1, 2, 3
\end{aligned}
$$

Example

For Asia, the model is the following:

. . . .

For Africa

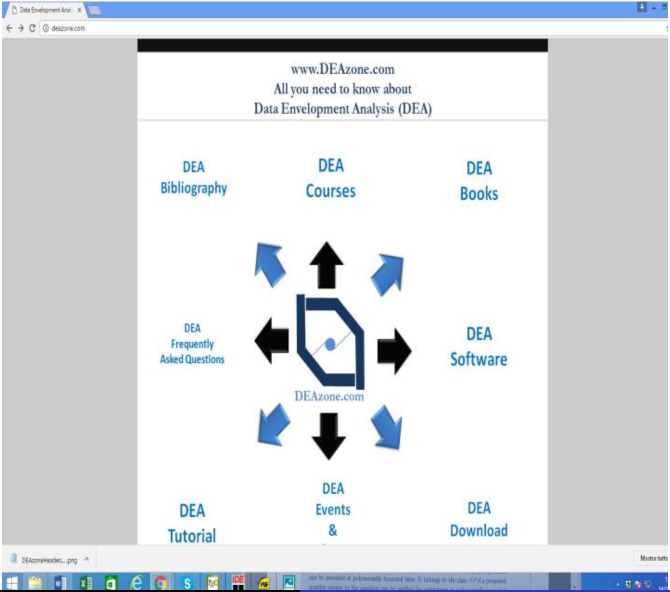
. . . .

For S. America

. . .

DMU	Efficiency
N. America	1
Europe	0.67
Asia	0.875
Africa	0.75
S. America	1

Additional materials



How efficiency can be improved

- How efficiency of inefficient DMUs can be improved ?
- Where do we get such information ?
- We may look at the dual formulation . . .

The Dual Formulations

We derive the dual formulation of the CCR input oriented. We rewrite the problem as

$$
\begin{aligned}
\max z = \; & \sum_{j=1}^{n} y_j^k t_j \\
& \sum_{i=1}^{m} x_i^k w_i = 1 \quad (\theta) \\
& \sum_{j=1}^{n} y_j^l t_j - \sum_{i=1}^{m} x_i^l w_i \le 0 \;\; \forall l \;\; (\lambda^l) \\
& -w_i \le -\delta \;\; \forall i \;\; (\sigma_i^-) \\
& -t_j \le -\delta \;\; \forall j \;\; (\sigma_j^+)
\end{aligned}
$$

Primal and Dual formulations

$$\max c^T x$$
$$Ax \leq b \qquad (P)$$
$$x \geq 0$$

$$\min b^T y$$
$$A^T y \geq c \qquad (D)$$
$$y \geq 0$$

The Dual CCR Input oriented

$$
\begin{aligned}
\min \quad & z = \theta - \delta(\sum_{i=1}^{m} \sigma_i^- + \sum_{j=1}^{n} \sigma_j^+) \\
& \sum_{l=1}^{K} y_j^l \lambda^l - \sigma_j^+ = y_j^k \quad \forall j \\
& \sum_{l=1}^{K} x_i^l \lambda^l + \sigma_i^- = \theta x_i^k \quad \forall i \\
& \lambda^l \geq 0 \ \ \forall l \\
& \sigma_j^+ \geq 0 \qquad \forall j \\
& \sigma_i^- \geq 0 \qquad \forall i \\
& \theta \qquad \text{free}
\end{aligned}
$$

The Dual CCR Input oriented: remarks

- From the duality theory, we know that the primal model has $(m + n + K + 1)$ constraints, whereas the dual formulation has $(m + n)$ constraints.
- Since the number K of DMU is typically far superior to $(n + m)$, the solution to the dual problem is typically easier
- From the theory of linear programming it is well known that the values of the dual variables correspond to the shadow prices of the primal problem.
- So variables λ^l represent the shadow prices associated with the constraints of efficiency.

The Dual CCR Input oriented: remarks

- If the corresponding constraint is satisfied for equality, then the corresponding shadow price will be positive, otherwise it will be 0.
- In addition, we know that if in the optimal solution of the primal formulation an efficiency constraint (different from the one assoiated with the evaluated DMU) is satisfied for equality, then the corresponding DMU will be inefficient.
- Therefore, positive shadow prices identify the Reference Set for each inefficient DMU.

The Dual CCR Input oriented: remarks

- The DEA technique provides a tool to improve the efficiency of inefficient DMU
- If DMU k is efficient then $z^* = 1$ ($\theta^* = 1$ $\sigma_j^{+*} = 0$, $\sigma_i^{-*} = 0$)
- If the DMU k is inefficient then $z^* < 1$ and some of the σ variables will be positive
- By loking at the dual constraints, we get information on how improve the efficiency

$$\sum_{l=1}^{K} y_j^l \lambda^{l*} = y_j^k + \sigma_j^{+*} \quad \forall j$$

$$\sum_{l=1}^{K} x_i^l \lambda^{l*} = \theta^* x_i^k - \sigma_i^{-*} \quad \forall i$$

The Dual CCR Input oriented: remarks

- The basic idea is to to create an ideal DMU that performs better than the analyzed one
- To this aim, we use the DMUs of the reference set by the variables λ^{I*} which take into the account the relative importance of each unit
- The new ideal DMU will use for each input i a quantity not greater than that used by the DMU k.
- Similarly, for each output j will produce a quantity not less than that produced by k.

The Dual CCR Input oriented: remarks

- To move an inefficient DMU on the efficient frontier we have to perform the following corrections
  - **Radial input reduction**

$$x_i^k \Longrightarrow \quad \theta^* x_i^k - \sigma_i^{-*}$$

  - **Output increase**

$$y_j^k \Longrightarrow \quad y_j^k + \sigma_j^{+*}$$

Example

How the efficiency of Europe can be improved ?
We may look at the solution of the dual problem

DMU	Efficiency
N. America	1
Europe	0.67
Asia	0.875
Africa	0.75
S. America	1

Example

- We record the values of λ

$$\lambda_1 = 0.158 \qquad \lambda_5 = 0.658$$

- They will indicate the weights used in creating the ideal DMU
- The efficiency of Europe can be improved either decreasing the inputs or increasing the outputs

Example

- The model suggests to increase the output levels

$$\sigma_2^+ = 1.263 \quad \sigma_3^+ = 2.737$$

- Thus the new output levels should be

$$y_1^2 = 8$$

$$y_2^2 = 2 + 1.263$$

$$y_3^2 = 9 + 2.737$$

- Moreover, the inputs should be radially reduced (all the σ_i^- are 0):

$$x_i^2 = \theta^* * x_i^2$$

The Dual CCR Output oriented

$$\max \quad z = \phi + \delta(\sum_{i=1}^{m} \sigma_i^- + \sum_{j=1}^{n} \sigma_j^+)$$

$$\sum_{l=1}^{K} y_j^l \lambda^l = \phi y_j^k + \sigma_j^+ \quad \forall j$$

$$\sum_{l=1}^{K} x_i^l \lambda^l = x_i^k - \sigma_i^- \quad \forall i$$

$$\lambda^l \geq 0 \quad \forall l$$

$$\sigma_j^+ \geq 0 \qquad \forall j$$

$$\sigma_i^- \geq 0 \qquad \forall i$$

$$\phi \qquad \text{free}$$

The question of constant return to scale

- The CCR models are based on the assumption of constant return to scale (CRS).
- It is assumed that an increase in the amount of inputs consumed would lead to a proportional increase of the amount of outputs produced.
- In real contexts, the assumption of the CRS is not always appropriate
- There exists production processes with variable return to scale (VRS), which are typically associated with economy (diseconomy) of scale
- Thus, in the VRS, the amount of outputs produced is deemed to increase more or less than proportionally than the increase in the inputs.
- The CRS version is more restrictive that the VRS and yields usually to a fewer number of efficient units and also lower efficiency scores among all DMUs.

The BCC model: primal formulation

- To account for VRS, we may define the efficiency as

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

- The resulting model, known as BBC model, is in the dual form

$$\begin{aligned}
\max z = \quad & \sum_{j=1}^{n} y_j^k t_j + \mu \\
& \sum_{i=1}^{m} x_i^k w_i = 1 \\
& \sum_{j=1}^{n} y_j^l t_j + \mu - \sum_{i=1}^{m} x_i^l w_i \leq 0 \quad l = 1, \dots, K \\
& w_i \geq \delta \quad i = 1, \dots, m \\
& t_j \geq \delta \quad j = 1, \dots, n \\
& \mu \quad \text{free}
\end{aligned}$$

The BCC model: dual formulation

$$\min \quad z = \theta - \delta(\sum_{i=1}^{m} \sigma_i^- + \sum_{j=1}^{n} \sigma_j^+)$$

$$\sum_{l=1}^{K} y_j^l \lambda^l - \sigma_j^+ = y_j^k \quad \forall j$$

$$\sum_{l=1}^{K} x_i^l \lambda^l + \sigma_i^- = \theta x_i^k \quad \forall i$$

$$\sum_{l=1}^{K} \lambda^l = 1$$

$$\lambda^l \geq 0 \quad \forall l$$

$$\sigma_j^+ \geq 0 \qquad \forall j$$

$$\sigma_i^- \geq 0 \qquad \forall i$$

$$\theta \qquad \text{free}$$

The BCC model: remarks

We observe that the feasible region of the BCC model is smaller than the one of the CCR formulation
Thus, a greater number of DMUs will be efficient