

# Data Envelopment Analysis

Kristian Wichmann

November 5, 2017

*Data envelopment analysis* - or DEA for short - is a technique for measuring relative efficiencies between a number of similar units. It is based on linear programming.

## 1 Efficiency - a problematic definition

The traditional measure of efficiency for a unit is simply:

$$\text{Efficiency} = \frac{\text{Output}}{\text{Input}} \quad (1.1)$$

However, this simple definition often falls short in practice: Often there will be several inputs and outputs, related to different resources, activities and factors. These may not all be equally important, which means that our definition above potentially runs into trouble.

### 1.1 Example

As an example, consider three factories (units):

- Unit 1 produces 100 items per day, and the inputs per item are 10 dollars for materials and 2 labour-hours.
- Unit 2 produces 80 items per day, and the inputs are 8 dollars for materials and 4 labour-hours.
- Unit 3 produces 120 items per day, and the inputs are 12 dollars for materials and 1.5 labour-hours.

This is summed up in table 1.1. It is not immediately clear how to rank the three units in terms of efficiency.

	Production ( $y_1$ )	Materials ( $x_1$ )	Labor ( $x_2$ )
Unit 1	100	10	2
Unit 2	80	8	4
Unit 3	120	12	1.5

Table 1: Data for the example.

## 2 Relative efficiency

One way out of this problem is to assign different *weights* to each input and output. Thus, a modified definition would be:

$$\text{Efficiency} = \frac{\text{Weighted output}}{\text{Weighted input}} \quad (2.1)$$

Or in slightly more mathematical terms, the efficiency of the  $j$ 'th unit can be expressed as:

$$E_j = \frac{\sum_{i=1}^{n_o} u_i y_{ij}}{\sum_{i=1}^{n_i} v_i x_{ij}} \quad (2.2)$$

Here,  $n_i$  and  $n_o$  are the numbers of inputs and outputs, respectively.  $x_{ij}$  and  $y_{ij}$  are the values of the  $i$ 'th input/output for the  $j$ 'th unit. And  $v_i$  and  $u_i$  are the chosen weights for the  $i$ 'th input/output. All weights are assumed to be non-negative.

Here, a universal set of weights  $\mathbf{u}$  and  $\mathbf{v}$  are used for all units. But different units might be operated in different ways - having different focuses and practices.

### 2.1 Example

For the three factories, this means that given the weights  $u_1$  corresponding to production,  $v_1$  corresponding to materials, and  $v_2$  corresponding to labor, the relative efficiencies are:

$$E_1 = \frac{100u_1}{10v_1 + 2v_2} \quad (2.3)$$

$$E_2 = \frac{80u_1}{8v_1 + 4v_2} \quad (2.4)$$

$$E_3 = \frac{120u_1}{12v_1 + 1.5v_2} \quad (2.5)$$

### 3 Determining weights according to a unit

Assuming differences between units are substantial, let's take the viewpoint, that each unit itself should be able to set its own weights to reflect its individual specialization.

So, given a specific unit  $j'$ , the weights should be chosen to maximize:

$$E_{j'} = \frac{\sum_{i=1}^{n_o} u_i y_{ij'}}{\sum_{i=1}^{n_i} v_i x_{ij'}} \quad (3.1)$$

This should be done subject to two criteria:

- All efficiencies should have a maximum of 1:

$$\forall j : E_j \leq 1 \quad (3.2)$$

- All weights should be non-negative:

$$\forall i : u_i \geq 0, \quad \forall i : v_i \geq 0 \quad (3.3)$$

#### 3.1 Linear programming

We wish to turn this into a linear programming problem. This is done by demanding the denominator in equation 3.1 to be constant (usually 1), and the maximizing the numerator:

- Maximize:  $\sum_{i=1}^{n_o} u_i y_{ij'}$ .
- Subject to:  $\sum_{i=1}^{n_i} v_i x_{ij'} = 1$ .

Each of the conditions in equation 3.2 can then be written:

$$\sum_{i=1}^{n_o} u_i y_{ij} - \sum_{i=1}^{n_i} v_i x_{ij} \leq 0 \quad (3.4)$$

#### 3.2 Example

Let say we want to determine the efficiency of factory 1. Then the quantity to be maximized is:

$$E_1 = \frac{100u_1}{10v_1 + 2v_2} \quad (3.5)$$

This is equivalent to maximizing  $100u_1$  subject to  $10v_1 + 2v_2 = 1$ . The upper limit to the efficiencies as expressed in equation 3.4 become:

$$100u_1 - 10v_1 - 2v_2 \leq 0 \quad (3.6)$$

$$80u_1 - 8v_1 - 4v_2 \leq 0 \quad (3.7)$$

$$120u_1 - 12v_1 - 1.5v_2 \leq 0 \quad (3.8)$$

And of course:

$$u_1 \geq 0, \quad v_1 \geq 0, \quad v_2 \geq 0 \quad (3.9)$$