Data Envelopment Analysis

Outline

1. Introduction
2. An example
3. Efficiency measure
4. Efficiency frontier
5. The CCR model
6. DEA model for measuring the competitiveness of nations
7. DEA model for bankruptcy prediction

1. Introduction

- Data Envelopment Analysis (DEA) is used to compare operating performance of a set of units such as companies, university departments, hospitals, bank branch offices, production plans or transportation systems.
- These units must be homogeneous for meaningful comparison.
- The performance measure is based on the results obtained by each unit (outputs) and on the resources utilized to achieve these results (inputs).
- For example, for bank branches, the outputs may consist of active bank accounts, checks cashed or loan raised. The inputs may be the number of cashiers, managers or rooms used at each branch.

2. An Example

- Consider a group of 3 hospitals.
- Assume each hospital has only the following two inputs:
  - Input 1 = capital (measured by the number of hospital beds)
  - Input 2 = labor (measured in thousands of labor hours used during one month)
- The outputs produced by each hospital are:
  - Output 1 = hundreds of patient-days during month for patients under age 14
  - Output 2 = hundreds of patient-days during month for patients between 14 and 65
  - Output 3 = hundreds of patient-days during month for patients over 65
- Inputs and outputs:

	Inputs		Outputs		
Hospital	1	2	1	2	3
I	5	14	9	4	16
II	8	15	5	7	10
III	7	12	4	9	13

An Example

- The efficiency of hospital i is defined to be

  Value of hospital i's outputs/cost of hospital i's inputs

- Let
  - t_r = price or value of one unit of output r
  - w_s = cost of one unit of input s

	Inputs		Outputs		
Hospital	**1**	**2**	**1**	**2**	**3**
I	5	14	9	4	16
II	8	15	5	7	10
III	7	12	4	9	13

- Efficiency of the three hospitals are then
  - Hospital 1 efficiency = $(9t_1 + 4t_2 + 16t_3)/(5w_1 + 14w_2)$
  - Hospital 2 efficiency = $(5t_1 + 7t_2 + 10t_3)/(8w_1 + 15w_2)$
  - Hospital 3 efficiency = $(4t_1 + 9t_2 + 13t_3)/(7w_1 + 12w_2)$

An Example

Four ideas are used in DEA to determine if a hospital is efficient:

1. No hospital can be more than 100% efficient:

   For hospital 1:

   $$(9t_1 + 4t_2 + 16t_3)/(5w_1 + 14w_2) \leq 1, \text{ or}$$

   $$9t_1 + 4t_2 + 16t_3 \leq 5w_1 + 14w_2, \quad \text{or}$$

   $$5w_1 + 14w_2 - 9t_1 - 4t_2 - 16t_3 \geq 0$$

2. Suppose we are interested in evaluating the efficiency of hospital i. We choose output prices (t_1,t_2, and t_3) and input cost (w_1 and w_2) that maximize efficiency.

   > If the efficiency of a hospital is 1, then it is efficient.
   >
   > If the efficiency is less than 1, than it is not efficient.

3. To simplify computation, input prices are scaled so that cost of hospital i's inputs equals to 1. For hospital 2, the additional constraint is $8w_1 + 15w_2 = 1$.

4. Ensure that each input cost and output price is strictly positive. Otherwise, DEA cannot detect inefficiency of an input with $w_j = 0$; or an output with $t_i = 0$.

An Example

Three linear programs:

Hospital 1 LP: max $z = \mathbf{9t_1 + 4t_2 + 16t_3}$

Subject to: $-9t_1 - 4t_2 - 16t_3 + 5w_1 + 14w_2 \geq 0$

$-5t_1 - 7t_2 - 10t_3 + 8w_1 + 15w_2 \geq 0$

$-4t_1 - 9t_2 - 13t_3 + 7w_1 + 12w_2 \geq 0$

$\mathbf{5w_1 + 14w_2 = 1}$

$t_1, t_2, t_3, w_1, w_2 \geq 0.0001$

Solution:

Hospital 1 efficiency = 1

Hospital 2 efficiency = 0.773

Hospital 3 efficiency = 1

Hospital 2 LP: max $z = \mathbf{5t_1 + 7t_2 + 10t_3}$

Subject to: $-9t_1 - 4t_2 - 16t_3 + 5w_1 + 14w_2 \geq 0$

$-5t_1 - 7t_2 - 10t_3 + 8w_1 + 15w_2 \geq 0$

$-4t_1 - 9t_2 - 13t_3 + 7w_1 + 12w_2 \geq 0$

$\mathbf{8w_1 + 15w_2 = 1}$

$t_1, t_2, t_3, w_1, w_2 \geq 0.0001$

Hospital 3 LP: max $z = \mathbf{4t_1 + 9t_2 + 13t_3}$

Subject to: $-9t_1 - 4t_2 - 16t_3 + 5w_1 + 14w_2 \geq 0$

$-5t_1 - 7t_2 - 10t_3 + 8w_1 + 15w_2 \geq 0$

$-4t_1 - 9t_2 - 13t_3 + 7w_1 + 12w_2 \geq 0$

$\mathbf{7w_1 + 12w_2 = 1}$

$t_1, t_2, t_3, w_1, w_2 \geq 0.0001$

An Example

SAS solution.

Hospital 2 LP: max z = $\mathbf{5t_1 + 7t_2 + 10t_3}$

Subject to: $-9t_1 - 4t_2 - 16t_3 + 5w_1 + 14w_2 \geq 0$

$-5t_1 - 7t_2 - 10t_3 + 8w_1 + 15w_2 \geq 0$

$-4t_1 - 9t_2 - 13t_3 + 7w_1 + 12w_2 \geq 0$

$\mathbf{8w_1 + 15w_2 = 1}$

$t_1, t_2, t_3, w_1, w_2 \geq 0.0001$

```
data;
input _id_ $ t1 t2 t3 w1 w2 _type_ $ _rhs_;
datalines;
 z            5   7   10  0  0  max     .
 hosp1            -9  -4  -16  5 14  ge    0
 hosp2            -5  -7  -10  8 15  ge    0
 hosp3            -4  -9  -13  7 12  ge    0
 input2            0    0    0  8 15  eq    1
 minvarvalue  0.0001 0.0001 0.0001 0.0001 0.0001 lowerbd .
 ;
proc lp;
run;
```

Output:

Variable Summary

Col	Name	Status	Variable Type	Price	Activity	Reduced Cost
1	t1	BASIC	LOWERBD	5	0.0798206	0
2	t2	BASIC	LOWERBD	7	0.0532753	0
3	t3	LOWBD	LOWERBD	10	0.0001	- 2.784615
4	w1	LOWBD	LOWERBD	0	0.0001	- 0.248205
5	w2	BASIC	LOWERBD	0	0.0666133	0
6	hosp1		SURPLUS	0	0	- 0.261538
7	hosp2	BASIC	SURPLUS	0	0.2269699	0
8	hosp3		SURPLUS	0	0	- 0.661538

Constraint Summary

Row	Name	Constraint Type	Rhs	Activity	Dual Activity
1	z	OBJECTVE	0	0.7730301	.
2	hosp1	GE	0	0	- 0.261538
3	hosp2	GE	0	0.2269699	0
4	hosp3	GE	0	0	- 0.661538
5	input2	EQ	1	1	0.7733333

An Example

SAS solution.

Output:

Variable Summary

Col	Name	Status	Type	Price	Activity	Reduced Cost
1	t1	BASIC	LOWERBD	5	0.0798206	0
2	t2	BASIC	LOWERBD	7	0.0532753	0
3	t3	LOWBD	LOWERBD	10	0.0001	- 2.784615
4	w1	LOWBD	LOWERBD	0	0.0001	- 0.248205
5	w2	BASIC	LOWERBD	0	0.0666133	0
6	hosp1		SURPLUS	0	0	- 0.261538
7	hosp2	BASIC	SURPLUS	0	0.2269699	0
8	hosp3		SURPLUS	0	0	- 0.661538

Constraint Summary

Row	Constraint Name	Type	Rhs	Activity	Dual Activity
1	z	OBJECTVE	0	0.7730301	.
2	hosp1	GE	0	0	- 0.261538
3	hosp2	GE	0	0.2269699	0
4	hosp3	GE	0	0	- 0.661538
5	input2	EQ	1	1	0.7733333

Optimal solution:

$Z = 5t_1 + 7t_2 + 10t_3 = 0.7730301$

$t_1 = 0.0798206$

$t_2 = 0.0532753$

$t_3 = 0.0001$

$w_1 = 0.0001$

$w_2 = 0.666133$

Dual variables:

$\lambda_1 = 0.261538$ *hospital1*

$\lambda_2 = 0$

$\lambda_3 = 0.661538$ *hospital3*

$u_4 = 0.773333 = \varepsilon$

(note the first 3 constraints are $\geq$ constraints)

An Example

R solution.

```
> library("lpSolve")
> library("linprog")
> c <- c(5,7,10,0,0)
> b <- c(0,0,0,1,0.0001,0.0001,0.0001,0.0001,0.0001)
> A <- rbind( c(-9,-4,-16,5,14),
+             c(-5,-7,-10,8,15),
+             c(-4,-9,-13,7,12),
+             c(0,0,0,8,15),
+             c(1,0,0,0,0), c(0,1,0,0,0), c(0,0,1,0,0), c(0,0,0,1,0), c(0,0,0,0,1))
> const_dir <- c(">=",">=",">=","=",">=",">=",">=",">=",">=")
> LP2 <- lp(direction="max",objective.in=c,const.mat=A,const.dir = const_dir,const.rhs=b)
> LP2
Success: the objective function is 0.7730301
```

An Example

R solution.

```
> LP2$solution
[1] 0.07982062 0.05327528 0.00010000 0.00010000 0.06661333
> LP2$objval
[1] 0.7730301
> lp(direction="max",objective.in=c,const.mat=A, const.dir
    =const_dir,const.rhs=b,compute.sens=TRUE)$duals
 [1] -0.2615385  0.0000000 -0.6615385  0.7733333
     0.0000000  0.0000000 -2.7846154 -0.2482051
     0.0000000  0.0000000
[11]  0.0000000  0.0000000  0.0000000  0.0000000
```

Note there are 14 dual variables:

1. There are 9 constraints $\Rightarrow$ 3 hospitals, input of hospital 2 = 1, 5 input and output variables at least 0.0001
2. Solver automatically assumes all variables to be non-negative $\Rightarrow$ additional 5 dual variables.

Optimal solution:

$Z = 5t_1 + 7t_2 + 10t_3 = 0.7730301$

$t_1 = 0.07982062$

$t_2 = 0.05327528$

$t_3 = 0.0001$

$w_1 = 0.0001$

$w_2 = 0.06661333$

Dual variables:

$\lambda_1 = 0.261538$ *hospital1*

$\lambda_2 = 0$

$\lambda_3 = 0.661538$ *hospital3*

$u_4 = 0.773333 = \varepsilon$

(note the first 3 constraints are $\geq$ constraints)

An Example

Why is Hospital 2 not efficient?

Hospital 2 LP: max z $= \mathbf{5t_1 + 7t_2 + \underline{10}t_3}$

Subject to: $-9t_1 - 4t_2 - 16t_3 + 5w_1 + 14w_2 \geq 0$

$-5t_1 - 7t_2 - 10t_3 + \mathbf{8w_1 + 15w_2} \geq 0$

$-4t_1 - 9t_2 - 13t_3 + 7w_1 + 12w_2 \geq 0$

$\mathbf{8w_1 + 15w_2 = 1}$

$t_1, t_2, t_3, w_1, w_2 \geq 0.0001$

- Create a composite hospital by combining 0.261538 of hospital 1 with 0.661538 of hospital 3.
- This composite hospital produces the same amount of output as hospital 2, but produces 12.785-10 = 2.785 more output 3 (patient days for over 65 patients).
- From the input vector, we see that the composite hospital uses less of each input than does hospital 2.

- Consider all the hospitals with nonzero dual prices.
- Hospital 1 dual price: 0.261538
- Hospital 3 dual price: 0.661538
- Average output vector with dual prices as weights:

$$0.261538 \begin{bmatrix} 9 \\ 4 \\ 16 \end{bmatrix} + 0.661538 \begin{bmatrix} 4 \\ 9 \\ 13 \end{bmatrix} = \begin{bmatrix} 5 \\ 7 \\ \underline{\mathbf{12.785}} \end{bmatrix}$$

- Average input vector:

$$0.261538 \begin{bmatrix} 5 \\ 14 \end{bmatrix} + 0.661538 \begin{bmatrix} 7 \\ 12 \end{bmatrix} = \begin{bmatrix} \mathbf{5.938} \\ \mathbf{11.6} \end{bmatrix}$$

$$\begin{bmatrix} 5 \\ 7 \\ 12.785 \end{bmatrix} > \begin{bmatrix} 5 \\ 7 \\ 10 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 5.938 \\ 11.6 \end{bmatrix} < \begin{bmatrix} 8 \\ 15 \end{bmatrix}$$

An Example

Why is Hospital 2 not efficient?

- The objective function value of $z = 0.7730$ for Hospital 2 LP implies that the more efficient composite hospital produces superior outputs by using at most 77.30% as much of each input:
  - Input 1 used by the composite hospital $< 0.7730 \times$ (input 1 used by hospital 2) $= 0.7730 \times 8 = 6.2$.
  - Input 2 used by the composite hospital $= 0.7730 \times$ (input 2 used by hospital 2) $= 0.77300 \times 15 = 11.6$.
  - More explanation on pages 24 - 28.

3. Efficiency measure

- The units being compared are called decision making units (DMUs)
- If the units produce a single output using a single input only, the efficiency of the i-th DMU is defined as

$$\theta_i = y_i/x_i$$

where y_i is the output value produced by DMU_i and x_i the input value used.

- If the units produce multiple outputs using various inputs, the efficiency is defined as the ratio between the weighted sum of the outputs and a weighted sum of the inputs.

For example, Hospital 2 efficiency = $(5t_1 + 7t_2 + 10t_3)/(8w_1 + 15w_2)$,

where the weights t_1, t_2, t_3 are associated with the outputs and w_1 and w_2 assigned to the inputs.

= [5(0.0798206) + 7(0.0532753) + 10(0.0001)]/1 = 0.7730301

Efficiency measure

- Example:
  - Hospital 1 efficiency = $(9t_1 + 4t_2 + 16t_3)/(5w_1 + 14w_2)$
  - Hospital 2 efficiency = $(5t_1 + 7t_2 + 10t_3)/(8w_1 + 15w_2)$
  - Hospital 3 efficiency = $(4t_1 + 9t_2 + 13t_3)/(7w_1 + 12w_2)$
- It is difficult to fix a single structure of weights that might be shared and accepted by all the 3 units/hospitals.
- In order to avoid possible objections raised by the units to a preset system of weights, DEA evaluates the efficiency of each unit through the weight system that is the best for the DMU itself, i.e.
  - the weight system that allows the efficiency value to be maximized for each DMU
  - one linear programming problem for each DMU to obtain its optimal weights.

4. Efficient frontier

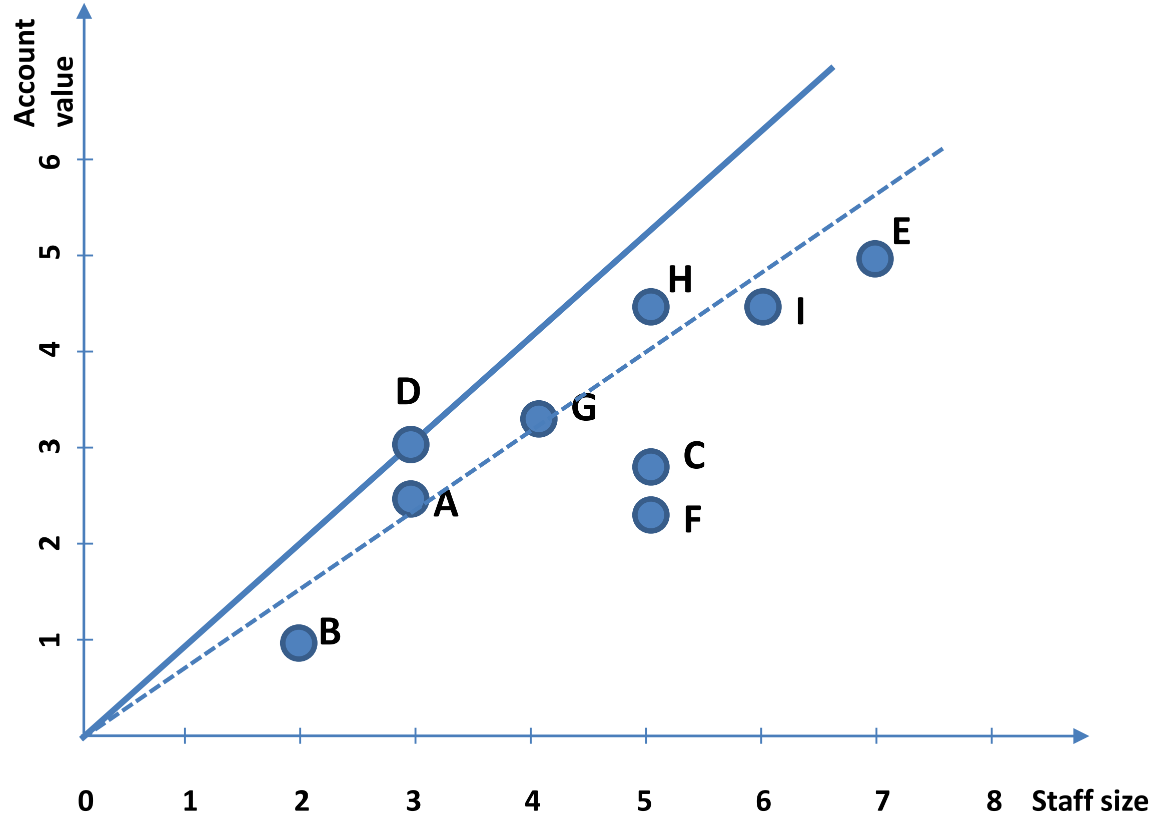
The efficient frontier:

- expresses the relationship between the inputs utilized and the outputs produced.
- also known as production function.
- indicates the maximum quantity of outputs that can be obtained from a given combination of inputs.
- expresses the minimum quantity of inputs that must be used to achieve a given output level.
- corresponds to technically efficient operating models.
- may be empirically obtained based on a set of observations that express the output level obtained by applying a specific combination of input production factors.

Efficient frontier

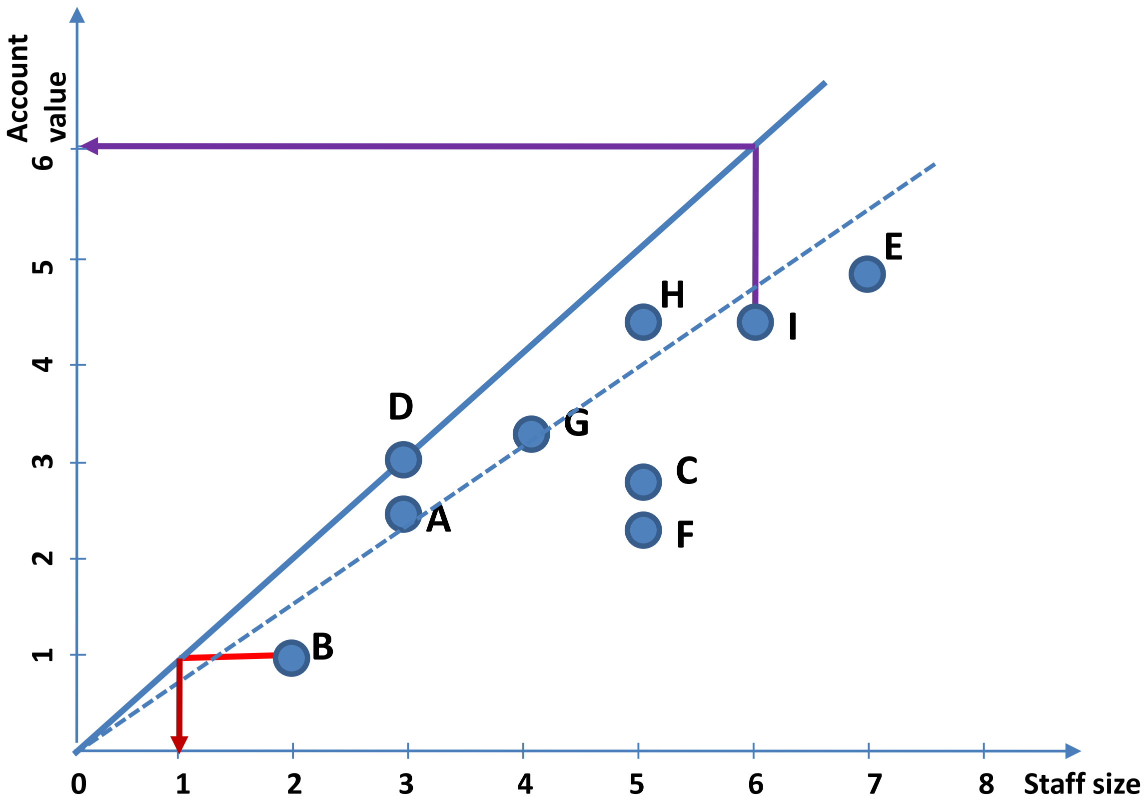
Example: Input and output values for nine bank branches.

Branch	Staff size	Account value	Efficiency
A	3	2.5	0.833
B	2	1.0	0.500
C	5	2.7	0.540
D	3	3.0	1.000
E	7	5.0	0.714
F	5	2.3	0.460
G	4	3.2	0.700
H	5	4.5	0.900
I	6	4.5	0.633



- Solid line: Line with maximum slope (= 1) is the efficient frontier
- Dotted line is a regression line, units that fall above the regression line may be deemed excellent and the degree of excellence of each unit could be expressed by its distance from the line.

Efficient frontier



- **Input oriented efficiency** is the ratio between ideal input quantity x* that should be used by the unit if it were efficient unit and the actually used quantity x_i:

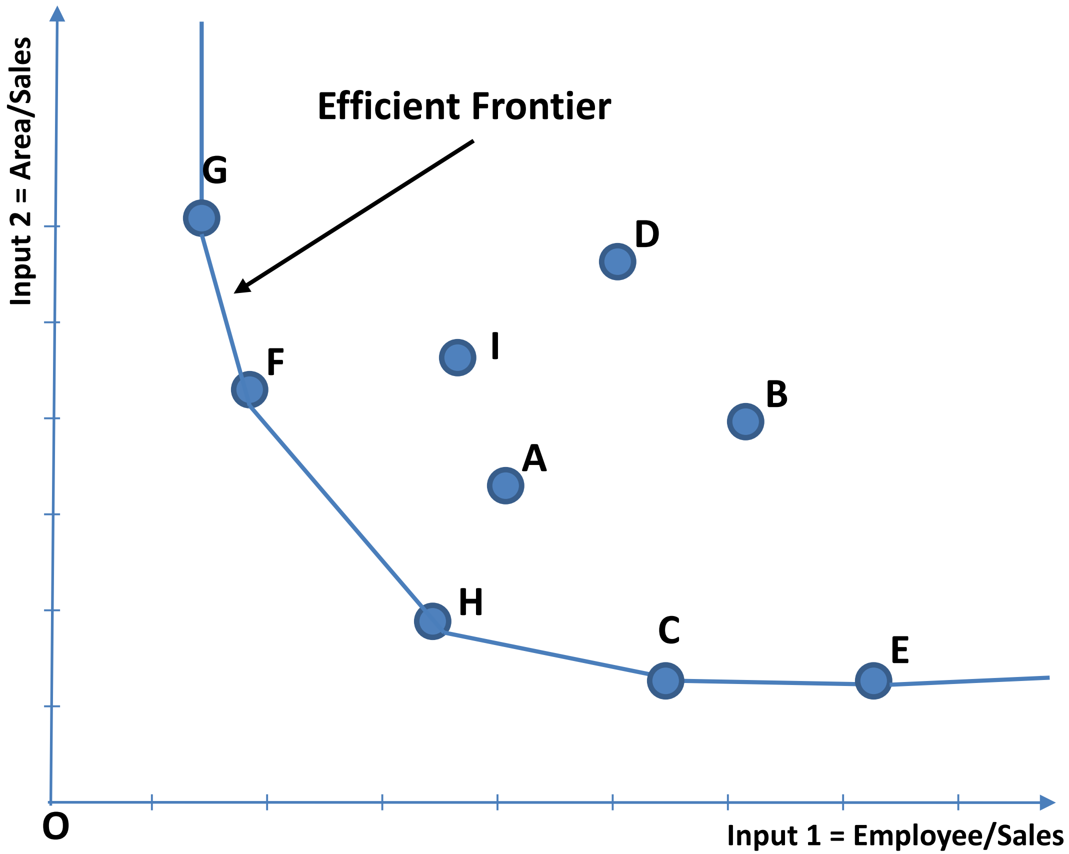
$$\theta^{Inp}_i = x^*/x_i$$

- **Output oriented efficiency** is the ratio between output quantity y_i actually produced by the unit and the ideal quantity y* that **it** should produce in conditions of efficiency:

$$\theta^{Out}_i = y_i/y^*$$

Efficient frontier

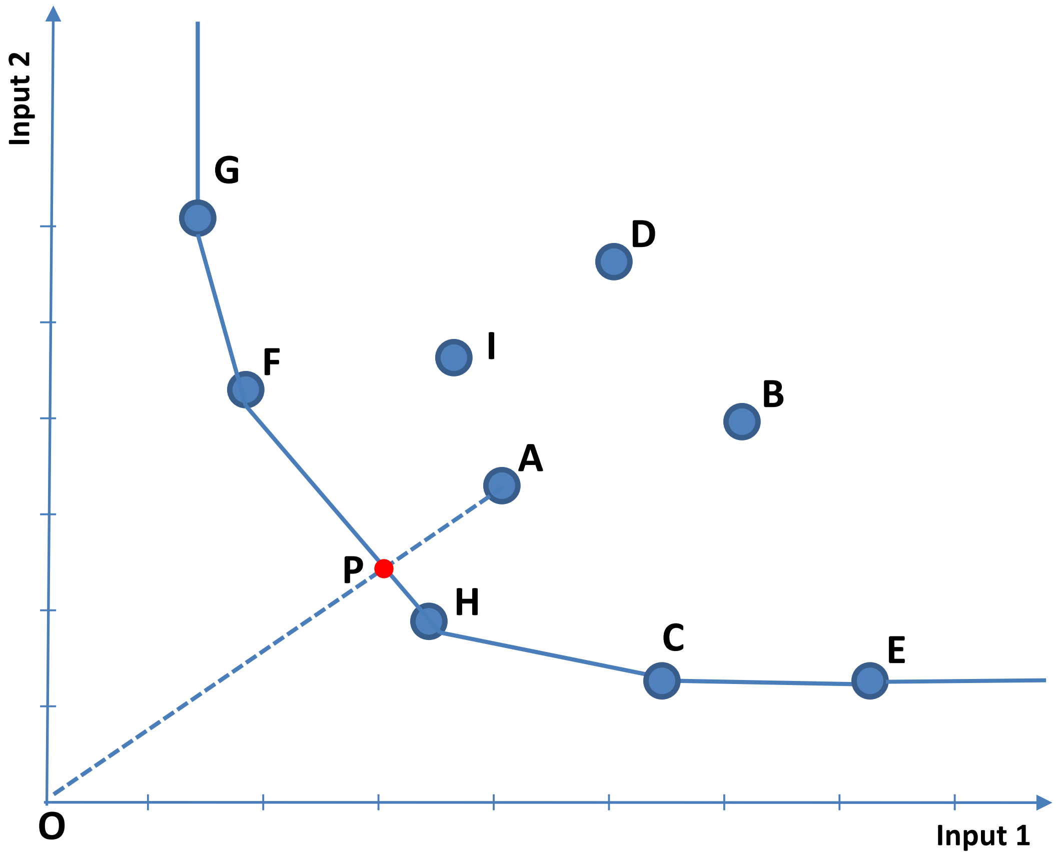
Efficiency frontier with two inputs and one output



- Assume that there are 9 supermarkets each with two inputs and one output.
- Input x_1 is the number of employees (unit: 10)
- Input x_2 is the floor area (1000 m2)
- Output y is the sales (100,000 $)
- Unitized sales to 1: normalize input values to values for getting 1 unit of sales.

Efficient frontier

Efficiency frontier with two inputs and one output



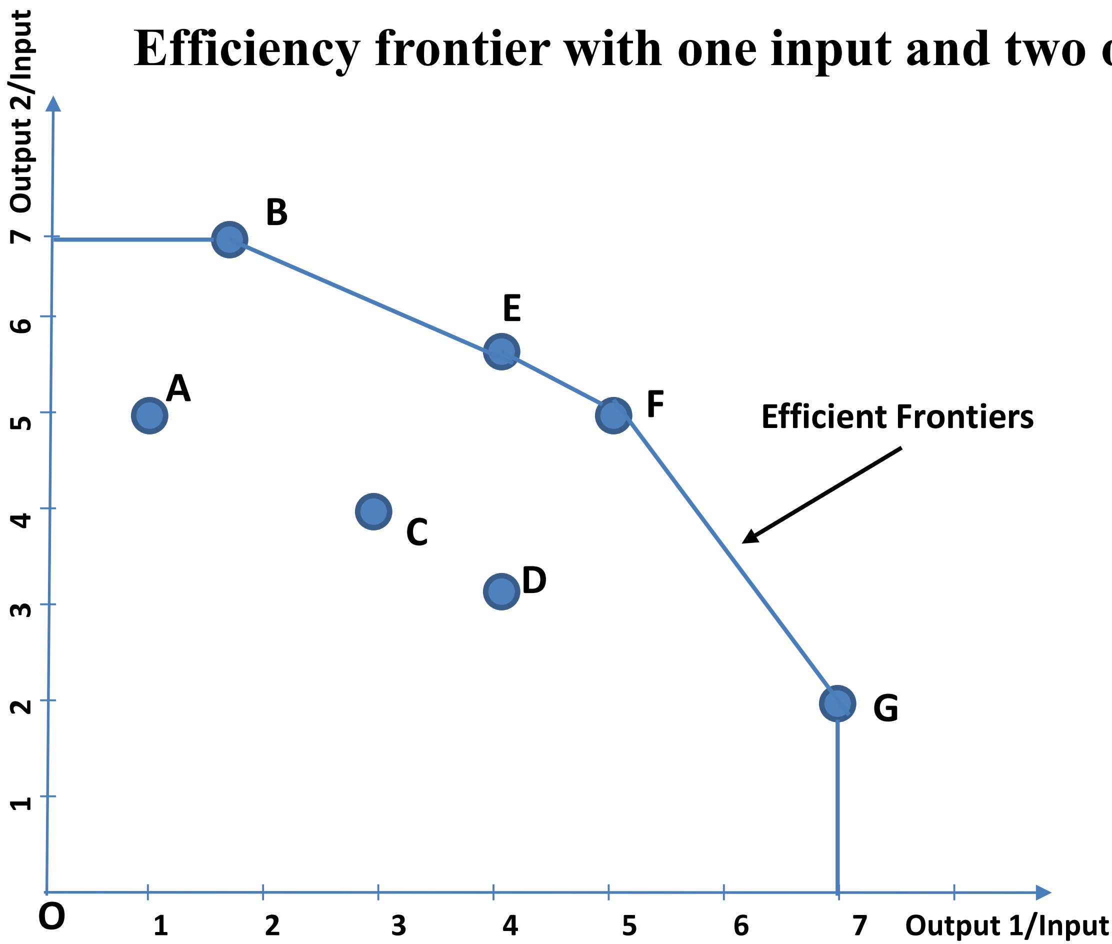
- The efficiency value of DMU_A is given by

$$\theta_A = OP/OA$$

- The inefficient unit may be made efficient by a displacement along segment OA that moves it onto the efficient frontier: decrease the quantity of both inputs while keeping the quantity of the output unchanged.

Efficient frontier

Efficiency frontier with one input and two outputs



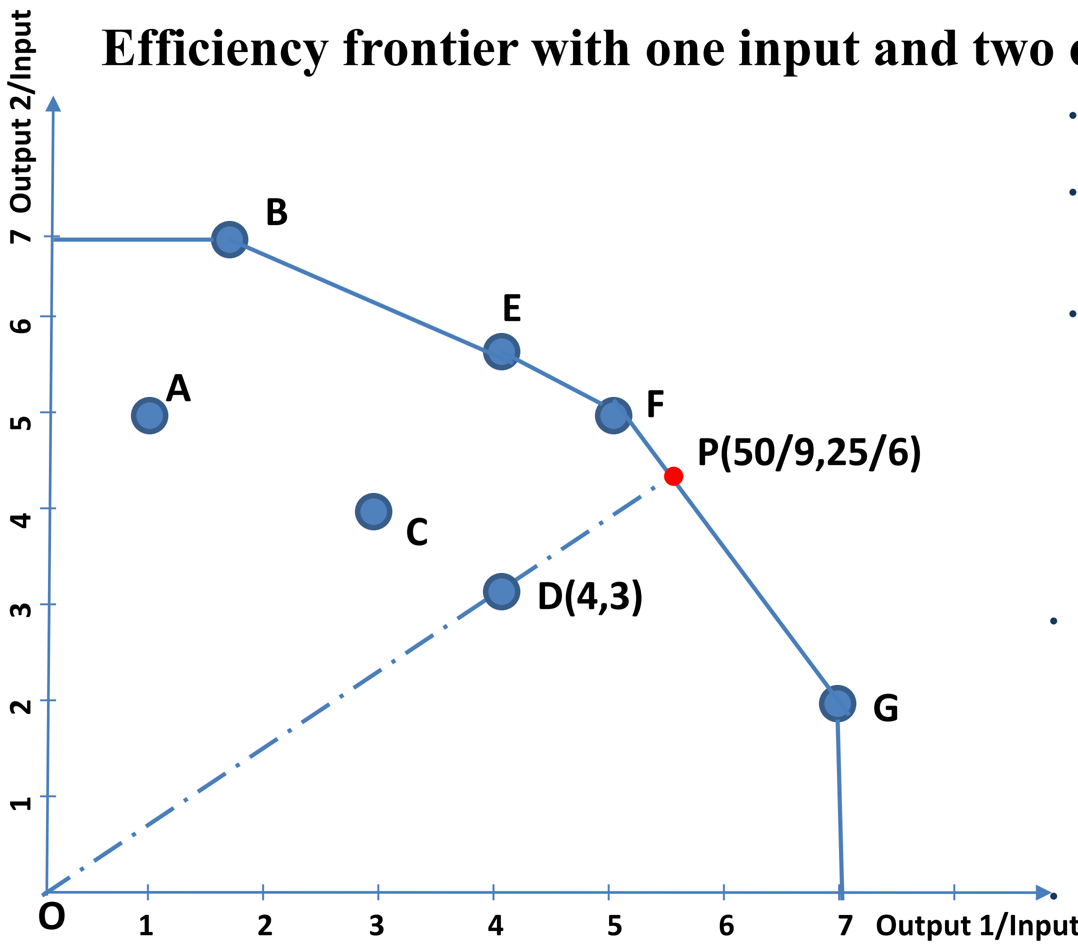
Consider the following data:

- Input: number of employees
- Output 1: the number of customers <u>per employee</u>
- Output 2: the amount of sales <u>per employee</u>

Store	Employees (x)	Customers (y_1)	Sales (y_2)
A	1	1	5
B	1	2	7
C	1	3	4
D	1	4	3
E	1	4	6
F	1	5	5
G	1	7	2

Efficient frontier

Efficiency frontier with one input and two outputs



0.72 = 4/(50/9) = 3/(25/6)

- Branches A, C and D are inefficient.
- Their efficiency can be evaluated by referring to the frontier lines.
- The efficiency of D is evaluated as

  d(O,D)/d(O,P)

  where d(O,D) is the distance from O to D:

  $d(O,D) = \text{sqrt}(4^2 + 3^2) = 5$

- P is on the line connecting F(5,5) and G(7,2), i.e.

  $y = -1.5x + 12.5$

  The slope of OP is equal to the slope of OD = 0.75

- We find that coordinates of P is (50/9, 25/6)

  and $d(O,P) = \text{sqrt}[(50/9)^2 + (25/6)^2] = 6.9444$

  $d(O,D)/d(O,P) = 5/6.9444 = 0.72$

- To be efficient, D would have to increase both of its outputs by 38.89% (1.3889 = 1/0.72) to P.

5. The CCR model

- What is the optimal system of weights for a DMU?
- The best known formulation is the Charnes-Cooper-Rhodes (CCR) model.
- For Hospital 2:

Hospital 2 LP: max $\vartheta = (5t_1 + 7t_2 + 10t_3)/(8w_1 + 15w_2)$

Subject to: $(9t_1 + 4t_2 + 16t_3)/(5w_1 + 14w_2) \leq 1$

$(5t_1 + 7t_2 + 10t_3)/(8w_1 + 15w_2) \leq 1$

$(4t_1 + 9t_2 + 13t_3)/(7w_1 + 12w_2) \leq 1$

$t_1, t_2, t_3, w_1, w_2 \geq 0$

- **Input oriented CCR model:** linearized the problem by requiring the weighted sum of the inputs equal to 1.

Hospital 2 LP: max $z = \mathbf{5t_1 + 7t_2 + 10t_3}$

Subject to: $9t_1 + 4t_2 + 16t_3 - 5w_1 - 14w_2 \leq 0$

$5t_1 + 7t_2 + 10t_3 - 8w_1 - 15w_2 \leq 0$

$4t_1 + 9t_2 + 13t_3 - 7w_1 - 12w_2 \leq 0$

$\mathbf{8w_1 + 15w_2 = 1}$

$t_1, t_2, t_3, w_1, w_2 \geq 0$

The CCR model

- Let z* be the maximum objective function value of the LP

Hospital 2 LP: max $z = \mathbf{5t_1 + 7t_2 + 10t_3}$

$$\text{Subject to: } 9t_1 + 4t_2 + 16t_3 - 5w_1 - 14w_2 \leq 0$$

$$5t_1 + 7t_2 + 10t_3 - 8w_1 - 15w_2 \leq 0$$

$$4t_1 + 9t_2 + 13t_3 - 7w_1 - 12w_2 \leq 0$$

$$\mathbf{8w_1 + 15w_2 = 1}$$

$$t_1, t_2, t_3, w_1, w_2 \geq 0$$

and (t*,w*) be the corresponding optimal solution.

- The DMU is said to be **efficient** if z* = 1 and if there exists at least one solution such that

$$t^* > 0 \text{ and } w^* > 0.$$

- Note: We have seen earlier that this DMU is <u>not</u> efficient.

The CCR model

- Primal LP:

Hospital 2 LP: max $z = \mathbf{5t_1 + 7t_2 + 10t_3}$

Subject to:
$$9t_1 + 4t_2 + 16t_3 - 5w_1 - 14w_2 \le 0$$
$$5t_1 + 7t_2 + 10t_3 - 8w_1 - 15w_2 \le 0$$
$$4t_1 + 9t_2 + 13t_3 - 7w_1 - 12w_2 \le 0$$
$$\mathbf{8w_1 + 15w_2 = 1}$$
$$t_1, t_2, t_3, w_1, w_2 \ge 0$$

Multipliers:

λ_1

λ_2

λ_3

ε

- Dual LP:

Hospital 2 DLP: min ε

Subject to:
$$5\lambda_1 + 8\lambda_2 + 7\lambda_3 - 8\varepsilon \le 0$$
$$14\lambda_1 + 15\lambda_2 + 12\lambda_3 - 15\varepsilon \le 0$$
$$9\lambda_1 + 5\lambda_2 + 4\lambda_3 - 5 \ge 0$$
$$4\lambda_1 + 7\lambda_2 + 9\lambda_3 - 7 \ge 0$$
$$16\lambda_1 + 10\lambda_2 + 13\lambda_3 - 10 \ge 0$$
$$\lambda_1,\ \lambda_2,\ \lambda_3 \ge 0$$

Note: take the derivative with respect to w_1, w_2, t_1, t_2, t_3 in this order to get DLP.

The CCR model

- Dual LP:

Hospital 2 DLP: min ε

Subject to: $5\lambda_1 + 8\lambda_2 + 7\lambda_3 - 8\varepsilon \le 0$

$14\lambda_1 + 15\lambda_2 + 12\lambda_3 - 15\varepsilon \le 0$

$9\lambda_1 + 5\lambda_2 + 4\lambda_3 - 5 \ge 0$

$4\lambda_1 + 7\lambda_2 + 9\lambda_3 - 7 \ge 0$

$16\lambda_1 + 10\lambda_2 + 13\lambda_3 - 10 \ge 0$

$\lambda_1,\ \lambda_2,\ \lambda_3 \ge 0$

The first two (input) constraints:

$$\begin{bmatrix} 5 \\ 14 \end{bmatrix} \lambda_1 + \begin{bmatrix} 8 \\ 15 \end{bmatrix} \lambda_2 + \begin{bmatrix} 7 \\ 12 \end{bmatrix} \lambda_3 \le \begin{bmatrix} 8 \\ 15 \end{bmatrix} \varepsilon$$

The next three (output) constraints:

$$\begin{bmatrix} 9 \\ 4 \\ 16 \end{bmatrix} \lambda_1 + \begin{bmatrix} 5 \\ 7 \\ 10 \end{bmatrix} \lambda_2 + \begin{bmatrix} 4 \\ 9 \\ 13 \end{bmatrix} \lambda_3 \ge \begin{bmatrix} 5 \\ 7 \\ 10 \end{bmatrix}$$

- If $\varepsilon^* < 1$, then this DMU lies below the efficient frontier. **Recall $\varepsilon^* = 0.7730$**
- In this case, it would be possible to use a quantity of inputs equal to a fraction of the quantity used by Hospital 2 and produce an output at least equal to the output it produces.
- This is achieved by creating a "composite hospital".

The CCR model

- What is the efficiency of the composite hospital?

Composite hospital LP: max $z = 5t_1 + 7t_2 + \mathbf{12.785}t_3$

Subject to:
$$9t_1 + 4t_2 + 16t_3 - 5w_1 - 14w_2 \le 0$$
$$5t_1 + 7t_2 + \mathbf{12.785}t_3 - \mathbf{5.938}w_1 - \mathbf{11.6}w_2 \le 0$$
$$4t_1 + 9t_2 + 13t_3 - 7w_1 - 12w_2 \le 0$$
$$\mathbf{5.938}w_1 + \mathbf{11.6}w_2 = 1$$
$$t_1, t_2, t_3, w_1, w_2 \ge 0$$

Recall the composite hospital is built by combining 0.261538 of hospital 1 with 0.661538 of hospital 3. (Page 12)

Eg.

12.785 = (0.261538)(16) + 0(10) + (0.661538)(13)

5.938 = (0.261538)(5) + (0.661538)(7)

11.6 = (0.261538)(14) + (0.661538)(12)

Composite Hospital DLP: min ε

Subject to:
$$5\lambda_1 + \mathbf{5.938}\lambda_2 + 7\lambda_3 - \mathbf{5.938}\varepsilon \le 0$$
$$14\lambda_1 + \mathbf{11.6}\lambda_2 + 12\lambda_3 - \mathbf{11.6}\varepsilon \le 0$$
$$9\lambda_1 + 5\lambda_2 + 4\lambda_3 - 5 \ge 0$$
$$4\lambda_1 + 7\lambda_2 + 9\lambda_3 - 7 \ge 0$$
$$16\lambda_1 + \mathbf{12.785}\lambda_2 + 13\lambda_3 - \mathbf{12.785} \ge 0$$
$$\lambda_1, \lambda_2, \lambda_3 \ge 0$$

- Hospital 1 dual price: 0.261538
- Hospital 3 dual price: 0.661538
- Average output vector with dual prices as weights:

$$0.261538\begin{bmatrix}9\\4\\16\end{bmatrix} + 0.661538\begin{bmatrix}4\\9\\13\end{bmatrix} = \begin{bmatrix}5\\7\\\underline{\mathbf{12.785}}\end{bmatrix}$$

- Average input vector:

$$0.261538\begin{bmatrix}5\\14\end{bmatrix} + 0.661538\begin{bmatrix}7\\12\end{bmatrix} = \begin{bmatrix}\mathbf{5.938}\\\mathbf{11.6}\end{bmatrix}$$

The CCR model

- What is the efficiency of the composite hospital?

Composite Hospital DLP: min ε

Subject to: $5\lambda_1 + \mathbf{5.938}\lambda_2 + 7\lambda_3 - \mathbf{5.938}\varepsilon \leq 0$

$14\lambda_1 + \mathbf{11.6}\lambda_2 + 12\lambda_3 - \mathbf{11.6}\varepsilon \leq 0$

$9\lambda_1 + 5\lambda_2 + 4\lambda_3 - 5 \geq 0$

$4\lambda_1 + 7\lambda_2 + 9\lambda_3 - 7 \geq 0$

$16\lambda_1 + \mathbf{12.785}\lambda_2 + 13\lambda_3 - \mathbf{12.785} \geq 0$

$\lambda_1,\ \lambda_2,\ \lambda_3 \geq 0$

The first two (input) constraints:

$$\begin{bmatrix} 5 \\ 14 \end{bmatrix} \lambda_1 + \begin{bmatrix} 5.938 \\ 11.6 \end{bmatrix} \lambda_2 + \begin{bmatrix} 7 \\ 12 \end{bmatrix} \lambda_3 \leq \begin{bmatrix} 5.938 \\ 11.6 \end{bmatrix} \varepsilon$$

The next three (output) constraints:

$$\begin{bmatrix} 9 \\ 4 \\ 16 \end{bmatrix} \lambda_1 + \begin{bmatrix} 5 \\ 7 \\ 12.785 \end{bmatrix} \lambda_2 + \begin{bmatrix} 4 \\ 9 \\ 13 \end{bmatrix} \lambda_3 \geq \begin{bmatrix} 5 \\ 7 \\ 12.785 \end{bmatrix}$$

- $\varepsilon = 1$, $\lambda_1 = 0.261538$, $\lambda_2 = 0$, $\lambda_3 = 0.661538$ is feasible for the dual problem.
- $\lambda_1 = \lambda_3 = 0$, $\lambda_2 = 1$ is also feasible for the dual problem with $\varepsilon = 1$.
- $\varepsilon = 1$ is the minimum value of DLP
- Hence, the composite hospital is efficient.

The CCR model: Another example

Consider the following data:

- Input: number of employees
- Output 1: the number of customers <u>per employee</u>
- Output 2: the amount of sales <u>per employee</u>

Store	Employees (x)	Customers (y_1)	Sales (y_2)
A	1	1	5
B	1	2	7
C	1	3	4
D	1	4	3
E	1	4	6
F	1	5	5
G	1	7	2

Store D LP: max $\vartheta = (4t_1 + 3t_2)/w_1$

Subject to: $(1t_1 + 5t_2)/w_1 \leq 1$

$(2t_1 + 7t_2)/w_1 \leq 1$

$(3t_1 + 4t_2)/w_1 \leq 1$

$(4t_1 + 3t_2)/w_1 \leq 1$

$(4t_1 + 6t_2)/w_1 \leq 1$

$(5t_1 + 5t_2)/w_1 \leq 1$

$(7t_1 + 2t_2)/w_1 \leq 1$

$t_1, t_2, \; w_1 \geq 0$

The CCR model: Another example

Store D LP: max $\vartheta = (4t_1 + 3t_2)/w_1$

Subject to: $(1t_1 + 5t_2)/w_1 \leq 1$

$(2t_1 + 7t_2)/w_1 \leq 1$

$(3t_1 + 4t_2)/w_1 \leq 1$

$(4t_1 + 3t_2)/w_1 \leq 1$

$(4t_1 + 6t_2)/w_1 \leq 1$

$(5t_1 + 5t_2)/w_1 \leq 1$

$(7t_1 + 2t_2)/w_1 \leq 1$

$t_1, t_2, w_1 \geq 0$

$w_1 = 1$ ⇒

Store D LP: max $z = 4t_1 + 3t_2$

Subject to: $1t_1 + 5t_2 \leq 1$ **A**

$2t_1 + 7t_2 \leq 1$ **B**

$3t_1 + 4t_2 \leq 1$ **C**

$4t_1 + 3t_2 \leq 1$ **D**

$4t_1 + 6t_2 \leq 1$ **E**

$5t_1 + 5t_2 \leq 1$ **F**

$7t_1 + 2t_2 \leq 1$ **G**

$t_1, t_2 \geq 0$

Solution of LP is:

$z = 0.72$,

$t_1 = 0.12$, $t_2 = 0.08$

$\lambda_A = \lambda_B = \lambda_C = \lambda_D = \lambda_E = 0$

$\lambda_F = 0.52 \quad \lambda_G = 0.20$

The CCR model: Another example

- Solution of LP is:

  z = 0.72,

  $t_1 = 0.12$, $t_2 = 0.08$

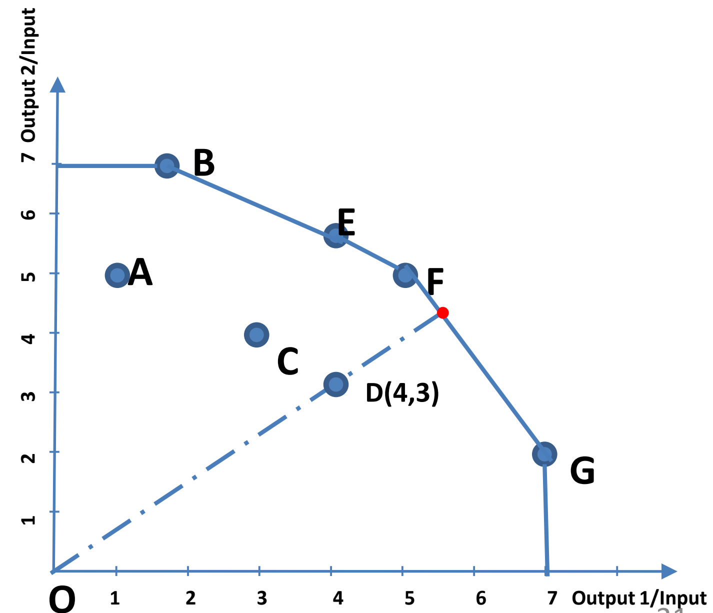
  $\lambda_A = \lambda_B = \lambda_C = \lambda_D = \lambda_E = 0$

  $\lambda_F = 0.52 \quad \lambda_G = 0.20$

- A new store that is a 'combination' of stores F and G is more efficient that store D. Its input is $\lambda_F\,(1) + \lambda_G\,(1) = 0.52 + 0.20 = 0.72$ and its output is

$$\lambda_F \begin{bmatrix} 5 \\ 5 \end{bmatrix} + \lambda_G \begin{bmatrix} 7 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$

- This new store has the same output as store D, but its input is only 0.72.



6. DEA model for measuring competitiveness of nations

- Reference: "The sustainable competitiveness of nations" by Sten Thore and Ruzanna Tarverdyan, Technological Forecasting & Social Change 106 (2016), 108-114.
- The competitiveness of a nation is measured as the ratio between the index of all policy goals achieved and a corresponding index of all policy instruments employed.
- **Output:** Y_r, $r = 1,2, \ldots s$ is a list of policy indicators.

  Goal index: $\mu_1 Y_1 + \mu_2 Y_2 + \mu_3 Y_3 + \ldots\ldots\ldots\ldots + \mu_s Y_s$
- **Input:** X_i, $i = 1,2, \ldots m$ is a list of competitiveness policy instruments.

  Policy instruments index: $\nu_1 X_1 + \nu_2 X_2 + \nu_3 X_3 + \ldots\ldots\ldots\ldots + \nu_m X_m$
- If there are n countries, the information can be summarized as (Y_{rj}, X_{ij}), $j = 1,2, \ldots. n$
- The effectiveness ratio for country $j = 0$ is then output/input:

$$(\mu_1 Y_{10} + \mu_2 Y_{20} + \mu_3 Y_{30} + \ldots\ldots + \mu_s Y_{s0}) / (\nu_1 X_{10} + \nu_2 X_{20} + \nu_3 X_{30} + \ldots\ldots + \nu_m X_{m0})$$

6. DEA model for measuring competitiveness of nations

- We would like to find $\mu_1, \mu_2, \mu_3, \ldots \mu_s$ and $v_1, v_2, v_3, \ldots v_m$ to <u>maximize</u>

$$(\mu_1 Y_{10} + \mu_2 Y_{20} + \mu_3 Y_{30} + \ldots\ldots + \mu_s Y_{s0}) / (v_1 X_{10} + v_2 X_{20} + v_3 X_{30} + \ldots\ldots + v_m X_{m0})$$

subject to:

$$(\mu_1 Y_{1j} + \mu_2 Y_{2j} + \mu_3 Y_{3j} + \ldots + \mu_s Y_{sj}) / (v_1 X_{1j} + v_2 X_{2j} + v_3 X_{3j} + \ldots + v_m X_{mj}) \le 1$$

for all j

$\mu_1 Y_{10} + \mu_2 Y_{20} + \mu_3 Y_{30} + \ldots\ldots + \mu_s Y_{s0} = 1$ (weighted index of all goals equal to 1)

- Linear program:

<u>minimize</u> $v_1 X_{10} + v_2 X_{20} + v_3 X_{30} + \ldots\ldots + v_m X_{m0}$

subject to

$$-\mu_1 Y_{1j} - \mu_2 Y_{2j} - \mu_3 Y_{3j} - \ldots - \mu_s Y_{sj} + v_1 X_{1j} + v_2 X_{2j} + v_3 X_{3j} + \ldots + v_m X_{mj} \ge 0$$

$j = 1, 2, \ldots n$

$\mu_1 Y_{10} + \mu_2 Y_{20} + \mu_3 Y_{30} + \ldots\ldots + \mu_s Y_{s0} = 1$

$\mu_r, v_i \ge 0$ for all $r = 1, 2, \ldots s$ and $i = 1, 2, \ldots m$

6. DEA model for measuring competitiveness of nations

- Eight goals:
  - Y_1 : GDP per person employed, 1990 USD equivalent
  - Y_2 : Equality of the income distribution
  - Y_3 : Youth employment
  - Y_4 : Access to sanitation
  - Y_5 : Access to improved drinking water
  - Y_6 : Agricultural water conservation
  - Y_7 : CO_2 release limitation
  - Y_8 : Forest cover conservation

s = 8

- Twelve pillars of input, X_1, X_2 … X_{12} are factors that promote the competitiveness of nations: health and primary education, higher education and training, financial market development, technology and innovation.

m = 12

- Number of countries in the study: 82

6. DEA model for measuring competitiveness of nations

- 54 countries out of 82 are efficient having maximal effectiveness rating of 1.
- Ten frontier countries serving as peers to the largest number of sub-frontier countries:

Country	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Number of refs
	GDP, $	1-Gini	Youth empl. %	Access sanitation, %	Access water, %	Water conserv, %	CO_2, %	Forest conserv. %	
Venezuela	36,900	76.3	79.4	100	100	5.0	2.9	98.3	25
Pakistan	8,500	70.0	92.3	48	91	0.7	3.1	98.9	17
Slovak Rep	33,500	74.7	66.0	100	100	4.8	3.0	97	14
Moldova	15,200	67.0	86.9	87	97	3.5	3.6	101	14
Armenia	29,300	68.7	64.6	91	100	1.3	3.3	100	9
US	68,400	61.1	84.5	100	99	2.1	2.5	95.9	9
Slovenia	36,900	76.3	79.4	100	100	5.0	2.9	99	8
Kazakhstan	25,400	71.0	96.1	97	93	1.0	1.7	99.8	8
Algeria	11,300	64.7	75.2	95	84	1.6	1.9	97.6	7
Uruguay	24,500	54.7	81.6	96	99	4.1	3.4	122	7

- Ten sub-frontier countries:

Country	Eff. score
Romania	94.8
Germany	94.4
Russia	94.2
Czech Rep.	94.1
Malaysia	91.6
Lithuania	91.3
New Zealand	90.5
Poland	89.3
Philippines	89.1
Indonesia	79.8

Country	Peers (with their λj weights)
Romania	Slovak Rep. (0.29), Moldova (0.25), Cameroon (0.21)
Germany	Venezuela (0.31), Austria (0.25), US (0.23)
Russia	Venezuela (0.54), Bolivia (0.46)
Czech Rep.	Slovak Rep. (0.59), Venezuela (0.22), Kazakhstan (0.16)
Malaysia	Venezuela (0.33), Guatemala (0.30), Kazakhstan (0.18)
Lithuania	Slovenia (0.28), Moldova(0.24), Armenia (0.23)
New Zealand	Norway (0.32), Moldova (0.28), Armenia (0.22)
Poland	Slovak Rep. (0.49), Venezuela (0.47), Moldova (0.08)
Philippines	Pakistan (0.36), Moldova (0.29), Venezuela (0.18)
Indonesia	Pakistan (0.56), Venezuela (0.39), Sri Lanka (0.10)

7. DEA model for bankruptcy prediction

- The article "DEA as a tool for bankruptcy assessment: A comparative study with logistic regression technique" by Premachandra et al. European Journal of Operational Research, Vol. 193 (2009) 412–424 describes how DEA can be used for classification.
- Linear Program:

$$
\begin{aligned}
\text{Max} \quad & \mathbf{e}^T s^- + \mathbf{e}^T s^+ \\
\text{subject to :} \quad & X\lambda + s^- = x_o \\
& Y\lambda - s^+ = y_o \\
& \mathbf{e}^T\lambda = 1 \\
& \lambda, s^-, s^+ \geq 0
\end{aligned}
$$

Where

- n is the number of DMUs, k is the number of inputs, m is the number of outputs
- **e** is a vector of all ones, $\mathbf{e} = (1,1, \ldots. 1)^T$
- X is k by n matrix of inputs, Y is m by n matrix of outputs
- x_o is the k-dimensional column vector of inputs of the o-th DMU
- y_o is the m-dimensional column vector of outputs of the o-th DMU
- s^- is the vector of input slacks
- s^+ is the vector of output slacks

DEA model for bankruptcy prediction

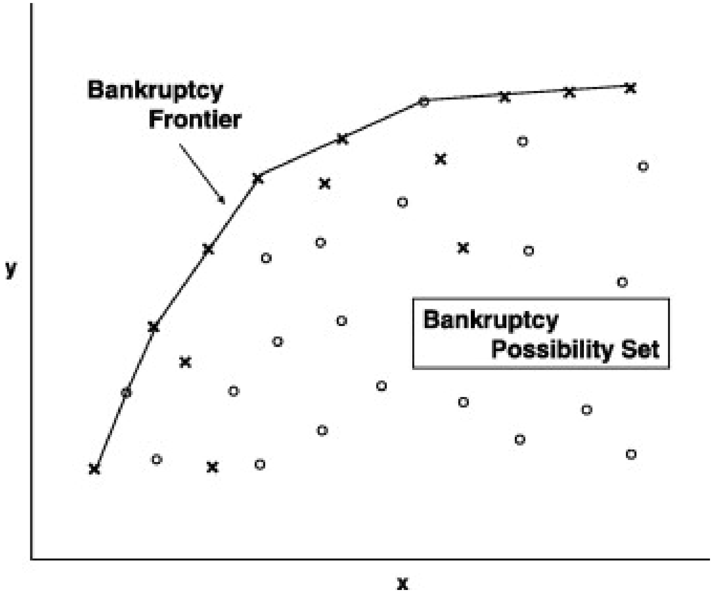
- **Seven input variables (k = 7):**
  - CFTA = cash flow/total assets.
  - NITA = net income/total assets.
  - WCTA = working capital/total assets.
  - CATA = current assets/total assets.
  - EBTA = earnings before interest and taxes/total assets.
  - EBIE = earnings before interest and taxes/interest expense.
  - MVCE = market value of equity/book value of common equity.
- **Two output variables (m = 2):**
  - Total debt/total assets (TDTA)
  - Current liabilities/total assets (CLTA)
- **Recall:** $X\lambda + s^- = x_o$

  $Y\lambda - s^+ = y_o$

  - Consider unit o with input x_o and output y_o, is it good to have positive slacks?

DEA model for bankruptcy prediction

Bankruptcy frontier:



Bankruptcy frontier and bankruptcy possibility set.

The symbol (○) indicates a non-default firm and the symbol (×) indicates a default firm.

If all slacks in the LP solution are zero,

then the firm is on the bankruptcy frontier.

Otherwise (at lease one slack is positive),

then the firm is not on the bankruptcy frontier.

Results:

	Appeared in Frontier	Did not appear in Frontier	Total
# Bankrupt (BR)	43	7	50
# Non-bankrupt (NBR)	294	616	910
Total	337	623	960

DEA model for bankruptcy prediction

Logistic regression model:

$$P_j = \frac{1}{1+e^{-Z_i}} = E\left(Y_i \middle| \begin{array}{l} \text{TDTA}_i, \text{CLTA}_i, \text{CFTA}_i, \text{NITA}_i, \text{WCTA}_i, \\ \text{CATA}_i, \text{EBTA}_i, \text{EBIE}_i \ \& \ \text{MVCE}_i \end{array}\right). \tag{2}$$

Here, Y_i is 1 if the firm is bankrupt and 0 otherwise,

$$\begin{aligned} Z_i = \beta_1 &+ \beta_2\text{TDTA} + \beta_3\text{CLTA} + \beta_4\text{CFTA} + \beta_5\text{NITA} \\ &+ \beta_6\text{WCTA} + \beta_7\text{CATA} + \beta_8\text{EBTA} + \beta_9\text{EBIE} \\ &+ \beta_{10}\text{MVCE} \end{aligned}$$

Result comparison:

On test data set, the DEA model significantly outperforms the LR model.

DEA accuracy: 74-86%

LR accuracy: 67%

References.

- W.L. Winston, Operations Research Applications and Algorithms, 4th Edition, Section 6.12
- C. Vercellis, Business Intelligence, Data Mining and Optimization for Decision Making, Chapter 15, Wiley.