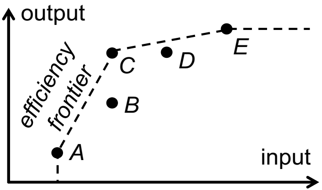
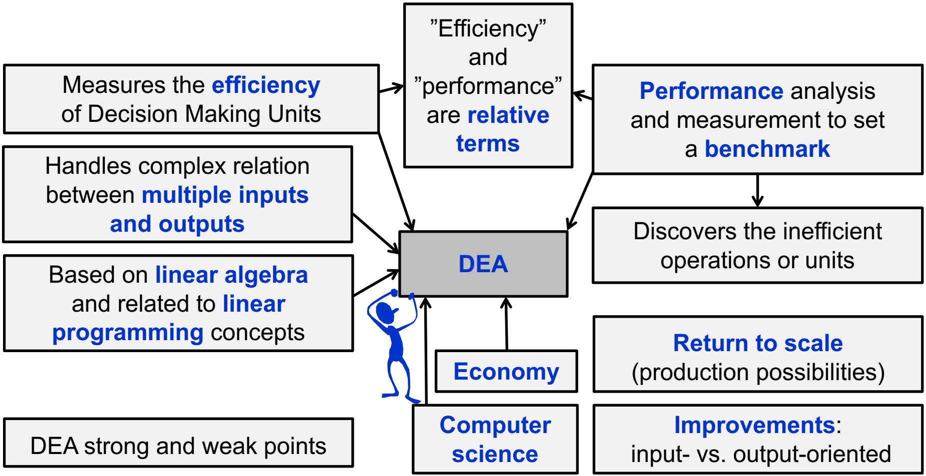


Data Envelopment Analysis: Measuring Efficiency of Decision Making Units

Miłosz Kadziński

Institute of Computing Science
Poznan University of Technology, Poland

Key concepts in DEA

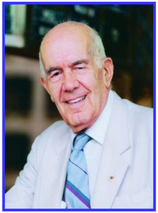


Seminal Papers on DEA

A. Charnes, W. Cooper, E. Rhodes, Measuring the efficiency of decision making units, *European Journal of Operational Research*, 2(6), 429-444,1978

- over 50000 citations (the most cited paper in EJOR)
- assumes **constant returns to scale** in production possibilities (an increase in the inputs leads to a proportional increase in outputs)

Abraham Charnes



William Cooper

Edwardo Rhodes

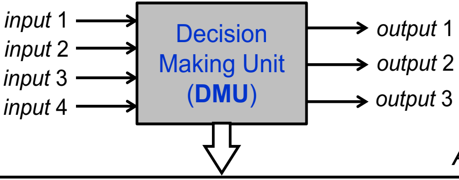
Rajiv Banker

R. Banker, A. Charnes, W. Cooper, Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis, *Management Science*, 30, 1078-1092,1984

- **variable return to scale** (efficiency depends on the scale of operations)

Decision Making Units

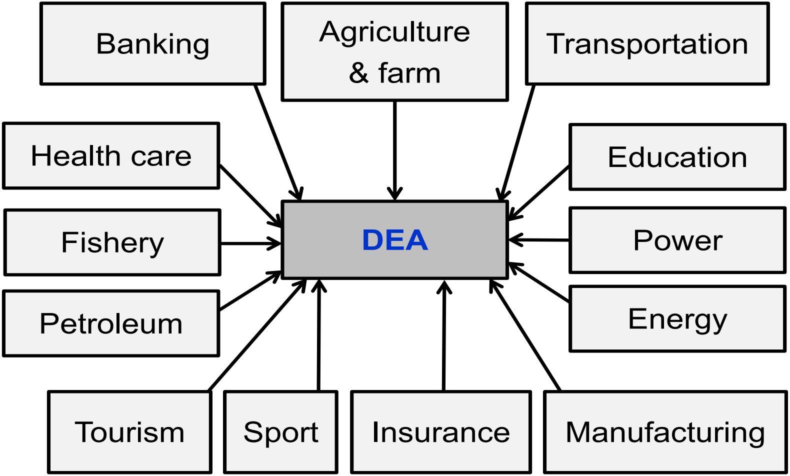
- We measure the performance and efficiency of **Decision Making Units**
- DMUs take some **input(s)** and produces some **output(s)**
- DMU is a very broad concept



Almost everything can be DMU

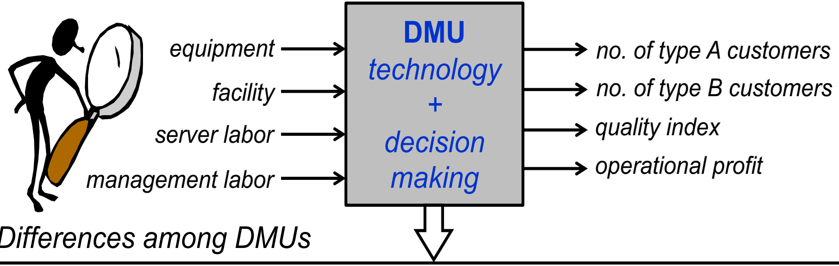
Factory's manufacturing units | Department store's | Airport
Business firms | Government agencies | Non-profit organization
University departments | Lecturers | Hospitals | Physicians

Main Application Areas of DEA



J. Liu, L. Lu, W-M. Lu, B. Lin, A survey of DEA applications, *Omega*, 41(5), 893-902,2013

Example of DMU: Sporting Goods Stores



Differences among DMUs

- Mix of customers served
- Availability and cost of inputs
- Configuration of production facilities
- Processes and practices used

Examples of Inputs and Outputs for Different DMUs

DMU	Inputs	Outputs
Manfacturing unit	raw materials, manpower floor space, energy	finished good (with added value)
Non-profit organization	volunteers' time donations, vehicles	impact on society
Bank	numbers of tellers and managers, computers	saving accounts, loan applications
Hospital	number of doctors, number of nurses	inpatients, outpatients
Airport	capacity of terminal and apron, catchment area	number of aviation operations, number of passengers
University	academic and non-academic staff, operating costs, area	students enrolled, completions, research income

Questions Answered by DEA

- How to **compare the efficiency** of diverse Decision Making Units?
- What are the **best-practice** and under-performing units?
- What are the **trade-offs among inputs and outputs**?
- Where the **improvement opportunities** / needs / requirements and how big they are?

DEA
TO THE RESCUE

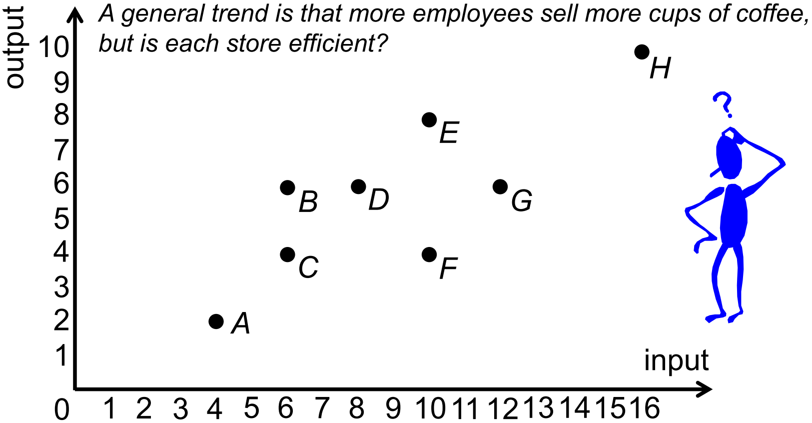
Example Scenario with One Input and One Output (1)

- Chain of **coffee shops** ”*Coffee and more*” in 8 locations (**A** to **H**)
- The owner wants to evaluate the **efficiency** of shops in one city
- He considers the **cups of coffee** sold (per day in thousands) as **output** and the **number is employees** in the store as **input**
- The owner wants to know *which stores are efficient*
- He is interested to benchmark the best store(s) so that he can *suggest improvements for the inefficient ones*

Store	DMU	A	B	C	D	E	F	G	H
Employees	**Input**	4	6	6	8	10	10	12	16
Cups of coffee	**Output**	2	6	4	6	8	4	6	10

Example Scenario with One Input and One Output (2)

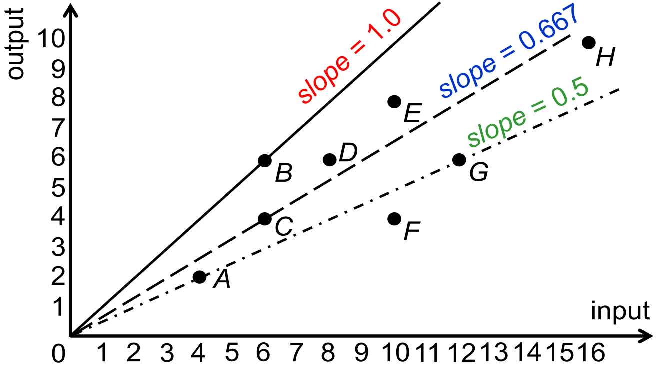
Store	DMU	A	B	C	D	E	F	G	H
Employees	**Input**	4	6	6	8	10	10	12	16
Cups of coffee	**Output**	2	6	4	6	8	4	6	10



Efficiency – First Idea

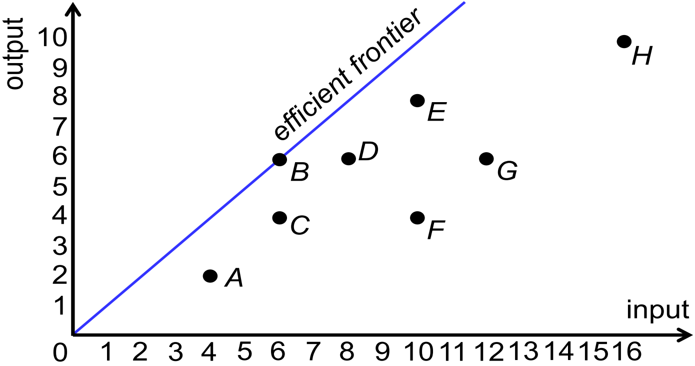
DMU	A	B	C	D	E	F	G	H
Input	4	6	6	8	10	10	12	16
Output	2	6	4	6	8	4	6	10
Ratio	0.5	1.0	0.667	0.75	0.80	0.4	0.5	0.625

The efficiency of each store is determined with the help of **ratio** *between the number of coffee cups sold and number of employees*



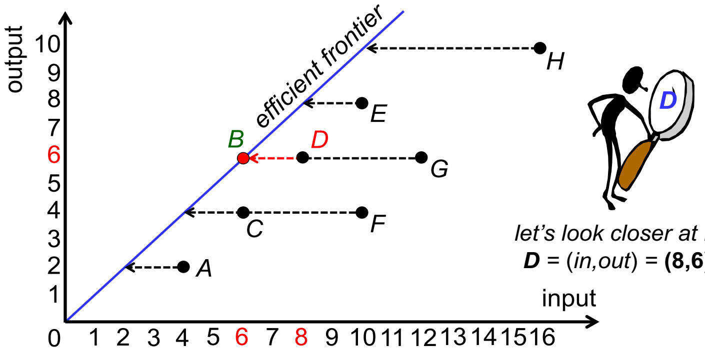
Efficient Frontier – First Idea

- The **frontier line** displays the performance of the best store in the comparison (stores **A**, **C-H** are "dominated" by store **B**)
- The efficiency of other stores can be measured by *the deviation of the respective points from the frontier line* (i.e., relative to the efficiency frontier)
- Efficient frontier serves as a **benchmark**



How You Can Improve? Focus on Inputs!

- **Input-oriented perspective**: excessive use of inputs by inefficient DMUs
- *How much less inputs an inefficient DMU should use in order to become efficient*?



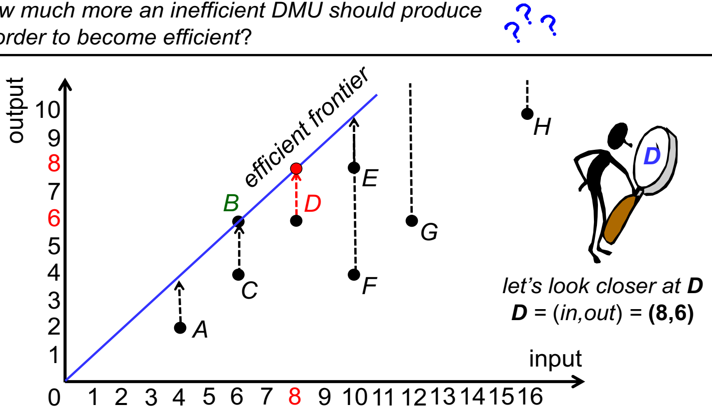
let's look closer at ***D***

D = (*in*,*out*) = **(8,6)**

- ***D*** should **produce its current output (6) with 2 units less of input** in order to reach the efficiency frontier (Δ*input* = 8 - 6 = 2)
- ***D*** should **produce its current output (6) with ¾ times units less of input** in order to reach the efficiency frontier (*desired input* = ¾ · 8 = 6)

How You Can Improve? Focus on Outputs!

- **Output-oriented perspective**: output gaps for inefficient DMUs
- *How much more an inefficient DMU should produce in order to become efficient*?

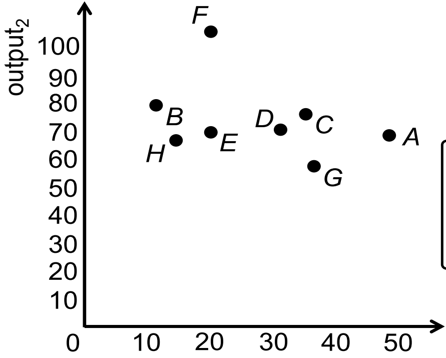


- ***D*** should **use its current input (8) to produce 2 units more of output** in order to reach the efficiency frontier (Δ*output* = 8 - 6 = 2)
- ***D*** should **use its current input (8) to produce 4/3 times units more of output** in order to reach the efficiency frontier (*desired output* = 4/3 · 6 = 8)

Example Scenario with One Input and Two Outputs

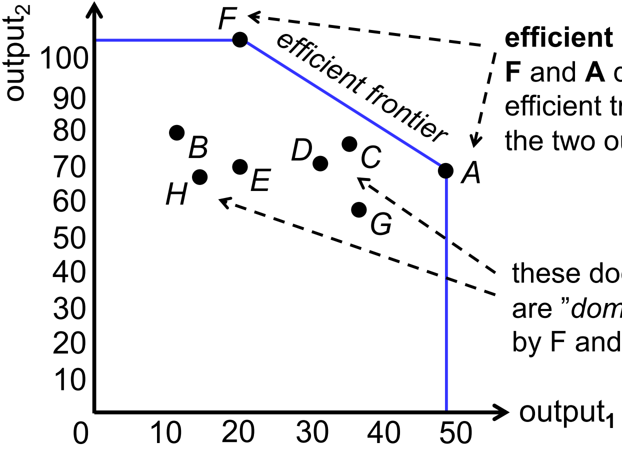
- Medical doctors (**A** to H) work at a hospital for the same **160 hours per month** (equal input), performing **exams** and **surgeries** (two outputs)
- Which ones are the most "productive"?

Doctor	DMU	A	B	C	D	E	F	G	H
Hours	**Input_1**	160	160	160	160	160	160	160	160
Exams	**Output_1**	48	12	35	31	20	20	36	15
Surgeries	**Output_2**	68	80	76	72	70	105	53	65



*There is some **"efficient" trade-off** between the numbers of surgeries and exams that any doctor can do in a month*

What is it?

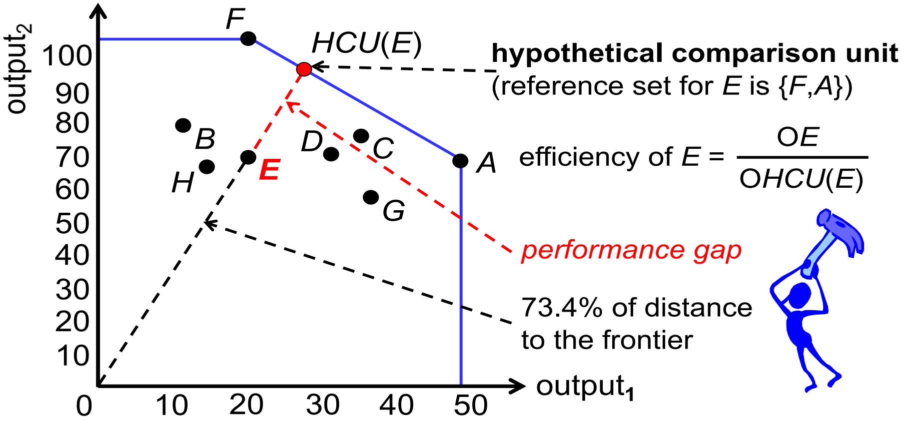


efficient doctors:
F and **A** define the most efficient trade-off between the two outputs

these doctors are "*dominated*" by F and A

"Pareto-Koopman efficiency" along the efficient frontier:
it is impossible to increase an output (or to decrease an input) without compensating decrease (increase) in other outputs (inputs)

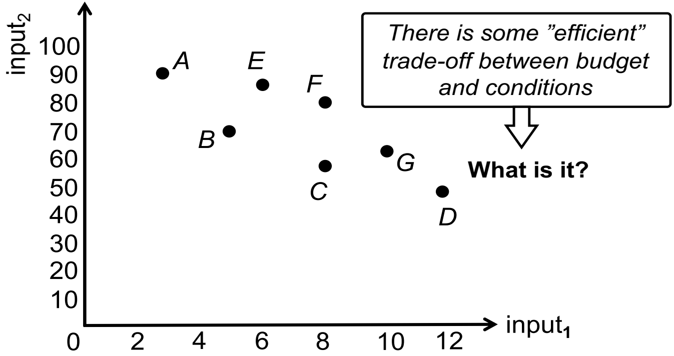
"Nearest" efficient DMUs define a **reference set** and linear combination of the reference set inputs and outputs of a **hypothetical comparison unit** (HCU)

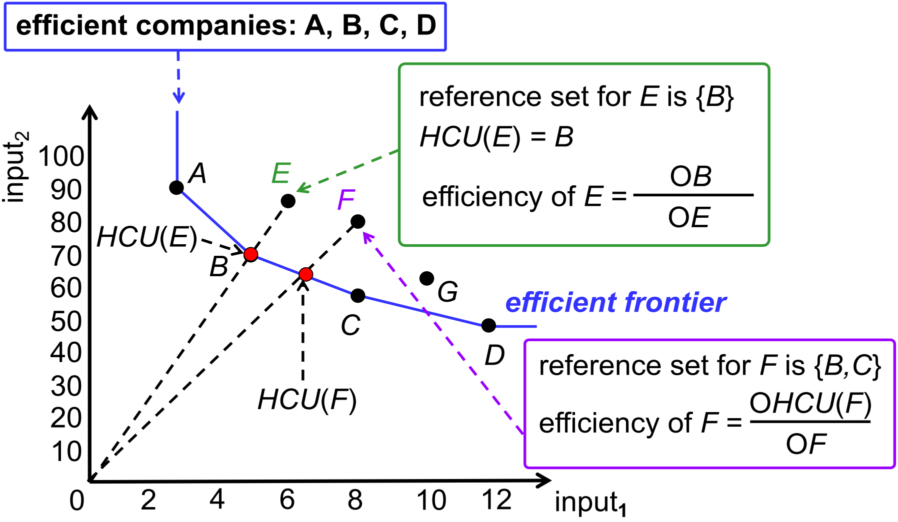


Example Scenario with Two Inputs and One Output (1)

- Different units (**A** to **G**) produce the same metal part having different budget and conditions
- Which ones are the most "productive"?

Unit	DMU	A	B	C	D	E	F	G
Budget	**Input$_1$**	3	5	8	12	6	8	10
Conditions	**Input$_2$**	90	70	55	50	84	80	60
Product	**Output$_1$**	1	1	1	1	1	1	1



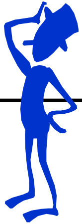

efficient companies: A, B, C, D
reference set for E is {B}
HCU(E) = B
efficiency of E = OB / OE
input2
100
90
80
70
60
50
40
30
20
10
0
2
4
6
8
10
12
input1
A
B
C
D
E
F
G
HCU(E)
HCU(F)
efficient frontier
reference set for F is {B,C}
efficiency of F = OHCU(F) / OF

Summary of Ideas Underlying DEA (so far)

- Input/output productivity is defined **relative to the efficient frontier**
- This frontier characterizes observed **efficient trade-offs** among inputs and outputs for a given set of DMUs
- **Efficiency** is defined as the relative distance to the frontier
- "Nearest point" on the frontier is the efficient HCU
- Differences in inputs and outputs between DMU and HCU correspond to **productivity "gaps"** (improvement potential)

⇓

How can we perform this analysis systematically?



How to Compare DMUs?

Financial approach

- Converting everything to monetary terms (€)

inputs in € → **DMU** → *outputs* in €

- Some inputs/outputs cannot be valued in € (non-profit)
- Profitability is not the same a operating effiiciency
- Improving operations calls for an understanding of technical efficiency (how much input is required to generate the output), not just overall profitability

Operating ratios

- *Examples*: labor hours per transaction, sales (€) per square meter
- Appropriate for highly standardized operations
- Do not reflect the varying mix of inputs/outputs of diverse operations

Measuring Efficiency

- Measuring the efficiency of DMUs is not always easy
- It can be easier for a company (e.g., measuring profits or added value)
  ... and more difficult for a non-profit organization

In its most basic form:

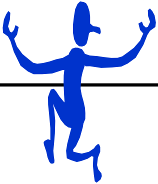
what is output?

$$\text{efficiency} = \frac{\text{output}}{\text{input}}$$

who decides what they are?

can we quantify them?

what is input?



The Ministry of Science has asked us to find universities with inefficient **management engineering departments**, since it considers shameful for such a department not to optimize its processes

Every department has a certain **number of professors** and, each year, produces a certain **number of graduates** and is able to secure a certain **amount of funding** for research

DMU	Input	Output$_1$	Output$_2$
University	**Professors**	**Graduates**	**Funds** (€)
PUT (DMU$_1$)	6	132	9600
AMU (DMU$_2$)	12	192	26400
UW (DMU$_3$)	10	190	21000
JU (DMU$_4$)	8	144	14400

Which university has an efficient / inefficient department?

Conical and Convex Combinations

	x_1	x_2
A	3	100
B	10	200

λ_A – weight (coefficient) of point A
λ_B – weight (coefficient) of point B

Conical combination is a linear combination of points where all coefficients are non-negative

$$\sum_{j=1,\dots,n} \lambda_j \cdot x_j = \lambda_1 \cdot x_1 + \lambda_2 \cdot x_2 + \dots + \lambda_n \cdot x_n \text{ for } \lambda_j \geq 0, \text{ for } j = 1, \dots, n$$

$\lambda_A = 0.3$ and $\lambda_B = 0.5$

x_1	x_2
$0.3 \cdot 3 + 0.5 \cdot 10 = 5.9$	$0.3 \cdot 100 + 0.5 \cdot 200 = 130$

Convex combination is a linear combination of points where all coefficients are non-negative and **sum to one**

$$\sum_{j=1,\dots,n} \lambda_j \cdot x_j = \lambda_1 \cdot x_1 + \lambda_2 \cdot x_2 + \dots + \lambda_n \cdot x_n \text{ for } \lambda_j \geq 0 \text{ and } \boldsymbol{\sum_{j=1,\dots,n} \lambda_j = 1}$$

$\lambda_A = 0.3$ and $\lambda_B = 0.7$

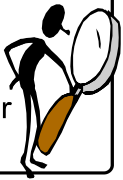
x_1	x_2
$0.3 \cdot 3 + 0.7 \cdot 10 = 8$	$0.3 \cdot 100 + 0.7 \cdot 200 = 170$

Running Example - Virtual (Combined) Unit (1)

DMU	Input	Output$_1$	Output$_2$
University	**Professors**	**Graduates**	**Funds (€)**
PUT (DMU$_1$)	6	132	9600
AMU (DMU$_2$)	12	192	26400
UW (DMU$_3$)	10	190	21000
JU (DMU$_4$)	8	144	14400
VU	**8**	**161**	**15300**

- Consider a department at a virtual university (**VU**) defined as the following conical / convex combination:

  VU = ½ · **PUT** + ½ · **UW**
- **VU** gives more output(s) with the same input of **JU**
- **JU** is inefficient, because some other DMUs or their combinations do better

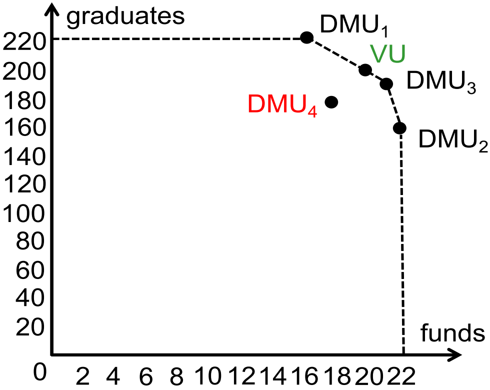


Running Example - Virtual (Combined) Unit (2)

University	Professors (in)	Graduates (out_1)	Funds (out_2)
PUT (DMU_1)	10 (6)	220 (132)	16000 (9600)
AMU (DMU_2)	10 (12)	160 (192)	22000 (26400)
UW (DMU_3)	10 (10)	190 (190)	21000 (21000)
JU (DMU_4)	**10 (8)**	**180 (144)**	**18000 (14400)**
VU	**10 (8)**	**201.25 (161)**	**19125 (15300)**

All inputs (number of professors) normalized to 10

We don't have to account for the input in the figure



JU is **inefficient under some assumptions**

- *outputs scale linearly with inputs*: if 10 professors can teach to 1000 students, can 20 professors teach to 2000 students?

- We do not consider economy of scale
- We do not consider operational complexity
- We assume **Constant Returns to Scale** (**CRS**)
- **All conical combinations of existing DMUs are realistic**

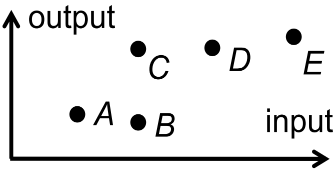
C C
C = R
R S

If a university produces Y output with X input,
every other university could be able to do the same
… but what if **JU** lectures are held in tents,
because Cracow has been hit by an earthquake?

Role of Efficiency Scores in DEA

Data Envelopment Analysis is a method to **assign a score** to each DMU, in order to measure its efficiency:

- **score = 1: DMU is efficient**
- **score < 1: DMU is inefficient**



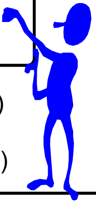
Efficiency is relative!

- DMU is **efficient** if it is not possible to obtain more output with less input, by combining other DMUs
- DMU is **inefficient** if it is possible to obtain more output with less input, by combining other DMUs

DEA: Notation

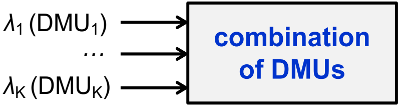
Data

- ***K*** operating DMUs, $k = 1, \ldots, K$ ($DMU_1, DMU_2, \ldots, DMU_K$)
- ***M*** inputs ($m = 1, \ldots, M$) ($input_1, input_2, \ldots, input_M$)
- $\boldsymbol{x}_{\mathbf{mk}}$ – observed level of $\mathbf{input_m}$ from $\mathbf{DMU_k}$
- ***N*** outputs ($n = 1, \ldots, N$) ($output_1, output_2, \ldots, output_N$)
- $\boldsymbol{y}_{\mathbf{nk}}$ – observed level of $\mathbf{output_n}$ from $\mathbf{DMU_k}$



Model variables

- $\boldsymbol{\lambda}_{\mathbf{k}}$ – weight (proportion) of DMU_k when constructing a combination
- $\boldsymbol{\theta}_{\mathbf{k}}$ – multiplier of inputs or outputs used to derive efficiency $\mathbf{E_k}$ of $\mathbf{DMU_k}$



On the Way to the Efficiency Model (1)

Data

- Let us first assume $\boldsymbol{\theta}_\mathbf{k} = \boldsymbol{E}_\mathbf{k}$
- Score $\boldsymbol{\theta}_\mathbf{k}$ should be **equal to 1** (=1) when the **DMU$_\mathbf{k}$** is **efficient** and less than 1 (<1) when it is **inefficient**
- Let us be more precise by seeing which values $\boldsymbol{\theta}_\mathbf{k}$ will be attained in case of inefficiency (<1)

For example, it is possible to find a linear combination of all DMUs that gives the same outputs as **DMU$_\mathbf{k}$**

$$\sum\nolimits_{j=1,\ldots,K} \lambda_j \cdot y_{nj} = y_{nk}, \qquad \boldsymbol{n} = 1, \ldots, N$$

using just three quarters of the inputs

$$\sum\nolimits_{j=1,\ldots,K} \lambda_j \cdot x_{mj} = \tfrac{3}{4} \cdot x_{mk}, \qquad \boldsymbol{m} = 1, \ldots, M$$

DMU$_k$

λ_1 (DMU$_1$) → comb. DMU
… →
λ_K (DMU$_K$) →

We would expect the (in)efficiency score of **DMU$_\mathbf{k}$** to be at most ¾: $\boldsymbol{\theta}_\mathbf{k} \leq$ ¾

On the Way to the Efficiency Model (2)

We say at most, because we could find another linear combination that gives the same outputs as **DMU$_k$**

$$\sum_{j=1,\dots,K} \lambda_j' \cdot y_{nj} = y_{nk}, \qquad \boldsymbol{n} = 1,\dots,N$$

using just half of the inputs

$$\sum_{j=1,\dots,K} \lambda_j' \cdot x_{mj} = \tfrac{1}{2} \cdot x_{mk}, \qquad \boldsymbol{m} = 1,\dots,M$$

DMU$_k$

λ_1 (DMU$_1$) →
…→
λ_K (DMU$_K$) →
comb. DMU

Then, we would know that efficiency score of **DMU$_k$** is at most ½: $\boldsymbol{\theta}_k \le \tfrac{1}{2}$

We would like to find the smallest $\boldsymbol{\theta}_k$ such that a combination produces the same (at least as good) output(s) of **DMU$_k$** using just a portion **DMU$_k$** of inputs

find the smallest … such that …

sounds familiar!

Input-oriented Combination-based CCR Model

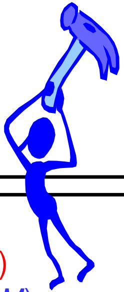
These inequalities suggest a first formulation of the DEA for **DMU$_k$** as a **linear problem**:

$$\min \boldsymbol{\theta}_k$$

$$\text{s.t.} \quad \sum_{j=1,\ldots,K} \lambda_j \cdot x_{mj} \leq \boldsymbol{\theta}_k \cdot x_{mk}, \qquad \boldsymbol{m} = 1, \ldots, M$$

$$\sum_{j=1,\ldots,K} \lambda_j \cdot y_{nj} \geq y_{nk}, \qquad \boldsymbol{n} = 1, \ldots, N$$

$$\boldsymbol{\theta}_k \geq 0; \; \lambda_k \geq 0, \; k = 1, \ldots, K$$



find the least $\boldsymbol{\theta}_k$ such that it is possible
to produce at least y_{nk} units of outputs ($\boldsymbol{n} = 1, \ldots, N$)
using no more than $\boldsymbol{\theta}_k \cdot x_{mk}$ units of inputs ($\boldsymbol{m} = 1, \ldots, M$),
for some linear combination of all the DMUs ($k = 1,\ldots, K$)

Reading Off Solution – Benchmarking

$\min \boldsymbol{\theta}_{\mathbf{k}}$

$$\text{s.t.} \quad \sum_{j=1,\ldots,K} \lambda_j \cdot x_{mj} \leq \boldsymbol{\theta}_{\mathbf{k}} \cdot x_{mk}, \qquad \boldsymbol{m} = 1, \ldots, M$$

$$\sum_{j=1,\ldots,K} \lambda_j \cdot y_{nj} \geq y_{nk}, \qquad \boldsymbol{n} = 1, \ldots, N$$

$$\boldsymbol{\theta}_{\mathbf{k}} \geq 0; \ \lambda_k \geq 0, \ k = 1, \ldots, K$$

note that $\boldsymbol{\theta}_k \leq 1$ because in the best case for **DMU**$_{\mathbf{k}}$ we can take $\boldsymbol{\lambda}_{\mathbf{k}}\mathbf{=1}$ and $\boldsymbol{\lambda}_{\mathbf{j}}\mathbf{=0}$ for all $j = 1, \ldots, K$ $(j \neq k)$

Under the assumptions made before, **DMU**$_{\mathbf{k}}$ could produce the same output, using just $\sum_{j=1,\ldots,K} \lambda_j \cdot x_{mj} = \boldsymbol{\theta}_{\mathbf{k}} \cdot x_{mk}$ input ($\boldsymbol{m} = 1, \ldots, M$), hence wasting:

$x_{mk} - \sum_{j=1,\ldots,K} \lambda_j \cdot x_{mj}$ units of input$_m$ for $m = 1, \ldots, M$

Input-oriented CCR Model – Example Formulation (1)

DMU_k	$input_1$	$output_1$	$output_2$
DMU_1	**6**	**132**	**9600**
DMU_2	12	192	26400
DMU_3	10	190	21000
DMU_4	8	144	14400

First model for **DMU_1** :

min $\boldsymbol{\theta_1}$

s.t.

$$6 \cdot \lambda_1 + 12 \cdot \lambda_2 + 10 \cdot \lambda_3 + 8 \cdot \lambda_4 \leq 6 \cdot \boldsymbol{\theta_1} \quad \Leftarrow \text{input}_1$$

$$132 \cdot \lambda_1 + 192 \cdot \lambda_2 + 190 \cdot \lambda_3 + 144 \cdot \lambda_4 \geq 132 \quad \Leftarrow \text{output}_1$$

$$\underset{\Uparrow \atop DMU_1}{9600 \cdot \lambda_1} + \underset{\Uparrow \atop DMU_2}{26400 \cdot \lambda_2} + \underset{\Uparrow \atop DMU_3}{21000 \cdot \lambda_3} + \underset{\Uparrow \atop DMU_4}{14400 \cdot \lambda_4} \geq \underset{\Uparrow \atop DMU_1}{9600} \quad \Leftarrow \text{output}_2$$

$\boldsymbol{\theta_1} \geq 0$; $\lambda_k \geq 0$, $k = 1, \ldots, 4$

Optimal solution: $\boldsymbol{\theta_1 = 1}$ for $\boldsymbol{\lambda_1 = 1}$ and $\boldsymbol{\lambda_2 = \lambda_3 = \lambda_4 = 0}$

Interpretation: **DMU_1** compares with itself; it is not possible to reduce the input of **DMU_1** using a combination of other DMUs

Conclusion: **DMU_1** is efficient

Input-oriented CCR Model – Example Formulation (2)

DMU_k	$input_1$	$output_1$	$output_2$
DMU_1	**6**	**132**	**9600**
DMU_2	12	192	26400
DMU_3	10	190	21000
DMU_4	8	144	14400

Repeat the same for each DMU

First model for **DMU_4** :

min $\boldsymbol{\theta_4}$

s.t.

$$6\cdot\lambda_1 + 12\cdot\lambda_2 + 10\cdot\lambda_3 + 8\cdot\lambda_4 \le 8\cdot\boldsymbol{\theta_4} \quad \Leftarrow \text{input}_1$$

$$132\cdot\lambda_1 + 192\cdot\lambda_2 + 190\cdot\lambda_3 + 144\cdot\lambda_4 \ge 144 \quad \Leftarrow \text{output}_1$$

$$9600\cdot\lambda_1 + 26400\cdot\lambda_2 + 21000\cdot\lambda_3 + 14400\cdot\lambda_4 \ge 14400 \quad \Leftarrow \text{output}_2$$

$$\Uparrow DMU_1 \quad \Uparrow DMU_2 \quad \Uparrow DMU_3 \quad \Uparrow DMU_4 \quad \Uparrow DMU_4$$

$\boldsymbol{\theta_4} \ge 0$; $\lambda_k \ge 0$, $k = 1,\dots, 4$

Optimal solution: $\boldsymbol{\theta_4}$ **= 0.911** for $\boldsymbol{\lambda_1}$ **= 0.3** and $\boldsymbol{\lambda_3}$ **= 0.55** and $\boldsymbol{\lambda_2 = \lambda_4}$ **= 0**

Conclusion: **DMU_4** is inefficient

$0.3\cdot DMU_1 + 0.55\cdot DMU_3 \Rightarrow$

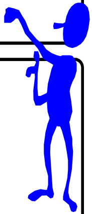
$input_1$	$output_1$	$output_2$
$8\cdot\boldsymbol{\theta_4} = 8\cdot\mathbf{0.911 = 7.29}$	144	14400

DMU_4 should reduce $input_1$ by 8 – 7.29 = 0.71 units to become efficient

Another Way of Looking at the Efficiency of DMUs (1)

Imagine we assign:

- weight v_m to **input**$_m$ ($m = 1, \ldots, M$) and weight u_n to **output**$_n$ ($n = 1, \ldots, N$)
- or, alternatively, unit price v_m to **input**$_m$ ($m = 1, \ldots, M$) and unit price u_n to **output**$_n$ ($n = 1, \ldots, N$) (these are just virtual values – one unit of any input or output does not really cost/ sells at this price)



Data

- K operating DMUs, $k = 1, \ldots, K$
- M inputs ($m = 1, \ldots, M$)
- x_{mk} – level of **input**$_m$ from **DMU**$_k$
- N outputs ($n = 1, \ldots, N$)
- y_{nk} – level of **output**$_n$ from **DMU**$_k$

Model variables

- v_m – weight of **input**$_m$ (unit price of **input**$_m$)
- u_n – weight of **output**$_n$ (unit price of **output**$_n$)
- θ_k – multiplier of inputs or outputs used to derive efficiency $\mathbf{E}_k$ of **DMU**$_k$

DEA: Notation (2)

Data

- x_{mk} – observed level of **input$_m$** from **DMU$_k$**
- v_m – weight of **input$_m$** (unit price of **input$_m$**)
- y_{nk} – observed level of **output$_n$** from **DMU$_k$**
- u_n – weight of **output$_n$** (unit price of **output$_n$**)

The **virtual input** of **DMU$_k$** / total cost of inputs for **DMU$_k$** is:

$$\sum_{m=1,\ldots,M} v_m \cdot x_{mk}$$

The **virtual output** of **DMU$_k$** / total gain from outputs of **DMU$_k$** is:

$$\sum_{n=1,\ldots,N} u_n \cdot y_{nk}$$

$$\textbf{efficiency} = \frac{\textbf{output}}{\textbf{input}} \Rightarrow \textbf{efficiency} = E_k = \frac{\sum_{n=1,\ldots,N} u_n \cdot y_{nk}}{\sum_{m=1,\ldots,M} v_m \cdot x_{mk}}$$

On the Way to Yet Another Efficiency Model

- Since weights/prices u_n and v_m are virtual, we can "play" with them and ask:

  what are the weights/prices u_n and v_m that maximize the efficiency of **DMU$_k$**?
- From equation:

$$E_k = \frac{\sum_{n=1,\dots,N} u_n \cdot y_{nk}}{\sum_{m=1,\dots,M} v_m \cdot x_{mk}}$$

  we see that, if we are free to "play" u_n and v_m as we want, then there is no limit: we could take u_n arbitrarily small and v_m arbitrarily big.
- Therefore **we impose a limit**:

$$E_k = \frac{\sum_{n=1,\dots,N} u_n \cdot y_{nk}}{\sum_{m=1,\dots,M} v_m \cdot x_{mk}} \le 1$$

- This choice is nice, as it gives a value between 0 and 1 for the efficiency, … but it is totally arbitrary: we could have said 2, 362, etc.

Input-oriented Efficiency-based CCR Model

- What are the weights/prices v_m and u_n that **maximize** the efficiency of **DMU$_k$** **such that** the efficiency for all DMUs is not greater than one
- **Maximize E_k such that $E_j \le 1$, for $j = 1, \ldots, K$**

$$\max \frac{\sum_{n=1,\ldots,N} u_n \cdot y_{nk}}{\sum_{m=1,\ldots,M} v_m \cdot x_{mk}}$$

$$\text{s.t.} \quad \frac{\sum_{n=1,\ldots,N} u_n \cdot y_{nj}}{\sum_{m=1,\ldots,M} v_m \cdot x_{mj}} \le 1 \text{ for } j = 1, \ldots, K$$

$$v_m \ge 0, \ m = 1, \ldots, M$$
$$u_n \ge 0, \ n = 1, \ldots, N$$

the Charnes-Cooper transformation

$$\max \sum_{n=1,\ldots,N} u_n \cdot y_{nk}$$

$$\text{s.t.} \quad \sum_{m=1,\ldots,M} v_m \cdot x_{mk} = 1$$

$$\sum_{n=1,\ldots,N} u_n \cdot y_{nj} \le \sum_{m=1,\ldots,M} v_m \cdot x_{mj}$$

$$v_m \ge 0, \ m = 1, \ldots, M$$
$$u_n \ge 0, \ n = 1, \ldots, N$$

Input-oriented CCR Model – Example Formulation

DMU_k	$Input_1$	$Output_1$	$Output_2$
DMU_1	**6**	**132**	**9600**
DMU_2	12	192	26400
DMU_3	10	190	21000
DMU_4	8	144	14400

Second model for **DMU_4** :

$$\max\ 144 \cdot u_1 + 14400 \cdot u_2$$

$$\begin{aligned}
\text{s.t.}\quad & 8 \cdot v_1 = 1 \\
& 132 \cdot u_1 + 9600 \cdot u_2 \leq 6 \cdot v_1 \\
& 192 \cdot u_1 + 26400 \cdot u_2 \leq 12 \cdot v_1 \\
& 190 \cdot u_1 + 21000 \cdot u_2 \leq 10 \cdot v_1 \\
& 144 \cdot u_1 + 14400 \cdot u_2 \leq 8 \cdot v_1 \\
& v_1, u_1, u_2 \geq 0
\end{aligned}$$

$v_1 = 0.125$
$u_1 = 0.0039557$
$u_2 = 0.0000237$
$E_4 = 0.911$
DMU_4 is inefficient

By definition the optimal solution gives us the efficiency of **DMU_k** :

$$\sum_{n=1,\ldots,N} \mathbf{u_n} \cdot \mathbf{y_{nk}} = \frac{\sum_{n=1,\ldots,N} \mathbf{u_n} \cdot \mathbf{y_{nk}}}{1} = \frac{\sum_{n=1,\ldots,N} \mathbf{u_n} \cdot \mathbf{y_{nk}}}{\sum_{m=1,\ldots,M} \mathbf{v_m} \cdot \mathbf{x_{mk}}} = E_k$$

Let's have a look at the two models we presented

$$\max \sum_{n=1,\dots,N} \boldsymbol{u}_{\mathbf{n}} \cdot \boldsymbol{y}_{\mathbf{nk}}$$

$$\text{s.t.} \quad \sum_{m=1,\dots,M} \boldsymbol{v}_{\mathbf{m}} \cdot \boldsymbol{x}_{\mathbf{mk}} = 1$$

$$- \sum_{m=1,\dots,M} \boldsymbol{v}_{\mathbf{m}} \cdot \boldsymbol{x}_{\mathbf{mj}} + \sum_{n=1,\dots,N} \boldsymbol{u}_{\mathbf{n}} \cdot \boldsymbol{y}_{\mathbf{nj}} \leq 0$$

$$\boldsymbol{v}_{\mathbf{m}} \geq 0, \ \boldsymbol{m} = 1, \dots, M$$

$$\boldsymbol{u}_{\mathbf{n}} \geq 0, \ \boldsymbol{n} = 1, \dots, N$$

$$\min \boldsymbol{\theta}_{\mathbf{k}}$$

$$\text{s.t.} \quad \boldsymbol{\theta}_{\mathbf{k}} \cdot x_{\mathrm{mk}} - \sum_{j=1,\dots,K} \lambda_{\mathrm{j}} \cdot x_{\mathrm{mj}} \geq 0, \quad \boldsymbol{m} = 1, \dots, M$$

$$\sum_{j=1,\dots,K} \lambda_{\mathrm{j}} \cdot y_{\mathrm{nj}} \geq y_{\mathrm{nk}}, \quad \boldsymbol{n} = 1, \dots, N$$

$$\theta_{\mathrm{k}} \geq 0; \ \lambda_{\mathrm{k}} \geq 0, \ k = 1, \dots, K$$

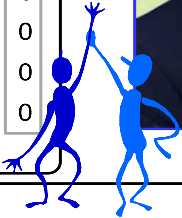
Any comments?

Duality of Combination- and Efficiency-based Models

First model for **DMU$_1$** :

$$\boxed{max}\quad \boxed{0\cdot v_1 + 132\cdot u_1 + 9600\cdot u_2}$$

$$\begin{array}{lrl}
s.t. & 6\cdot v_1 & = 1 \\
 & -6\cdot v_1 + 132\cdot u_1 + 9600\cdot u_2 & \leq 0 \\
 & -12\cdot v_1 + 192\cdot u_1 + 26400\cdot u_2 & \leq 0 \\
 & -10\cdot v_1 + 190\cdot u_1 + 21000\cdot u_2 & \leq 0 \\
 & -8\cdot v_1 + 144\cdot u_1 + 14400\cdot u_2 & \leq 0 \\
 & v_1, u_1, u_2 \geq 0 &
\end{array}$$



It is the dual!

Second model for **DMU$_1$** :

$$\boxed{min}\quad \boxed{1\cdot\theta_1 + 0\cdot\lambda_1 + 0\cdot\lambda_2 + 0\cdot\lambda_3 + 0\cdot\lambda_4}$$

$$\begin{array}{lrl}
s.t. & 6\cdot\theta_1 - 6\cdot\lambda_1 - 12\cdot\lambda_2 - 10\cdot\lambda_3 - 8\cdot\lambda_4 & \geq 0 \\
 & 132\cdot\lambda_1 + 192\cdot\lambda_2 + 190\cdot\lambda_3 + 144\cdot\lambda_4 & \geq 132 \\
 & 9600\cdot\lambda_1 + 26400\cdot\lambda_2 + 21000\cdot\lambda_3 + 14400\cdot\lambda_4 & \geq 9600 \\
 & \theta_1 \geq 0;\ \lambda_k \geq 0,\ k = 1,\ldots, 4 &
\end{array}$$

Output-Oriented Efficiency Analysis (1)

Find the greatest $\boldsymbol{\theta}_k$ such that it is possible
to use no more than x_{mk} units of inputs ($\boldsymbol{m} = 1, \ldots, M$),
producing at least $\boldsymbol{\theta}_k \cdot y_{nk}$ units of outputs ($\boldsymbol{n} = 1, \ldots, N$)
for some linear combination of all the DMUs ($k = 1, \ldots, K$)

max $\boldsymbol{\theta}_k$

s.t. $\sum_{j=1,\ldots,K} \lambda_j \cdot x_{mj} \leq x_{mk}, \quad \boldsymbol{m} = 1, \ldots, M$

$\sum_{j=1,\ldots,K} \lambda_j \cdot y_{nj} \geq \boldsymbol{\theta}_k \cdot y_{nk}, \quad \boldsymbol{n} = 1, \ldots, N$

$\boldsymbol{\theta}_k \geq 0; \ \lambda_k \geq 0, \ k = 1, \ldots, K$

note that $\boldsymbol{\theta}_k \geq 1$ because
in the best case for **DMU$_k$**
we can take $\boldsymbol{\lambda}_k\mathbf{=1}$ and
$\boldsymbol{\lambda}_j\mathbf{=0}$ for all $j = 1, \ldots, K$ ($j \neq k$)

Under the assumptions made before, with the same input **DMU$_k$** could produce $\sum_{j=1,\ldots,K} \lambda_j \cdot y_{nj} = \boldsymbol{\theta}_k \cdot y_{nk}$ output ($\boldsymbol{n} = 1, \ldots, N$), hence missing:

$\sum_{j=1,\ldots,K} \lambda_j \cdot y_{nj} - y_{nk}$ units of output$_n$ for $\boldsymbol{n} = 1, \ldots, N$

Output-oriented CCR Model – Example Formulation (1)

DMU_k	$input_1$	$output_1$	$output_2$
DMU_1	**6**	**132**	**9600**
DMU_2	12	192	26400
DMU_3	10	190	21000
DMU_4	8	144	14400

First model for $\mathbf{DMU_4}$:

$$\max \ \boldsymbol{\theta_4}$$

$$\text{s.t.} \quad 6\cdot\lambda_1 + 12\cdot\lambda_2 + 10\cdot\lambda_3 + 8\cdot\lambda_4 \le 8 \quad \Leftarrow \text{input}_1$$

$$132\cdot\lambda_1 + 192\cdot\lambda_2 + 190\cdot\lambda_3 + 144\cdot\lambda_4 \ge 144\cdot\boldsymbol{\theta_4} \quad \Leftarrow \text{output}_1$$

$$9600\cdot\lambda_1 + 26400\cdot\lambda_2 + 21000\cdot\lambda_3 + 14400\cdot\lambda_4 \ge 14400\cdot\boldsymbol{\theta_4} \quad \Leftarrow \text{output}_2$$

$$\boldsymbol{\theta_4} \ge 0; \ \lambda_k \ge 0, \ k = 1,\dots, 4$$

Optimal solution: $\boldsymbol{\theta_4} = \mathbf{1.097}$ for $\boldsymbol{\lambda_1} = \mathbf{0.33}$ and $\boldsymbol{\lambda_3} = \mathbf{0.60}$ and $\boldsymbol{\lambda_2} = \boldsymbol{\lambda_4} = \mathbf{0}$

Conclusion: $\mathbf{DMU_4}$ is inefficient (**its efficiency is $1/\boldsymbol{\theta_4}$ = 1/1.097=0.911**)

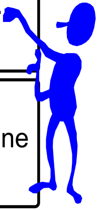
$0.33\cdot DMU_1 + 0.6\cdot DMU_3 \Rightarrow$

$input_1$	$output_1$	$output_2$
8	144·1.097=158	14400·1.097=15800

$\mathbf{DMU_4}$ should increase its $output_1$ by 158 – 144 = 14 units and $output_2$ by 15800 – 14400 = 1400 units to become efficient

Output-Oriented Efficiency Analysis (2)

$$\textbf{efficiency} = \frac{\textbf{input}}{\textbf{output}} \;\Rightarrow\; \textbf{efficiency} = E_k = \frac{\sum_{m=1,\dots,M} v_m \cdot x_{mk}}{\sum_{n=1,\dots,N} u_n \cdot y_{nk}}$$



- What are the weights/prices u_m and v_n that **minimize** the efficiency of **DMU$_k$ such that** the efficiencies for all DMUs are not less than one
- **Minimize E_k such that $E_j \geq 1$, for $j = 1, \dots, K$**

$$\min \frac{\sum_{m=1,\dots,M} v_m \cdot x_{mk}}{\sum_{n=1,\dots,N} u_n \cdot y_{nk}}$$

$$\text{s.t.} \quad \frac{\sum_{m=1,\dots,M} v_m \cdot x_{mj}}{\sum_{n=1,\dots,N} u_n \cdot y_{nj}} \geq 1 \text{ for } j = 1, \dots, K$$

$$v_m \geq 0, \; m = 1, \dots, M$$
$$u_n \geq 0, \; n = 1, \dots, N$$

$$\min \sum_{m=1,\dots,M} v_m \cdot x_{mk}$$

$$\text{s.t.} \quad \sum_{n=1,\dots,N} u_n \cdot y_{nk} = 1$$

$$\sum_{m=1,\dots,M} v_m \cdot x_{mj} \geq \sum_{n=1,\dots,N} u_n \cdot y_{nj}$$

$$v_m \geq 0, \; m = 1, \dots, M$$
$$u_n \geq 0, \; n = 1, \dots, N$$

DMU_k	$Input_1$	$Output_1$	$Output_2$
DMU_1	**6**	**132**	**9600**
DMU_2	12	192	26400
DMU_3	10	190	21000
DMU_4	8	144	14400

Second model for **DMU_4** :

$$\begin{aligned} & \min \; 8 \cdot v_1 \\ \text{s.t.} \; & 144 \cdot u_1 + 14400 \cdot u_2 = 1 \\ & 132 \cdot u_1 + 9600 \cdot u_2 \leq 6 \cdot v_1 \\ & 192 \cdot u_1 + 26400 \cdot u_2 \leq 12 \cdot v_1 \\ & 190 \cdot u_1 + 21000 \cdot u_2 \leq 10 \cdot v_1 \\ & 144 \cdot u_1 + 14400 \cdot u_2 \leq 8 \cdot v_1 \\ & v_1, u_1, u_2 \geq 0 \end{aligned}$$

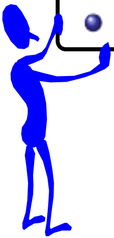
$v_1 = 0.1371527$
$u_1 = 0.000434027$
$u_2 = 0.000026041$
$8 \cdot v_1 = 1.097$
DMU_4 is inefficient

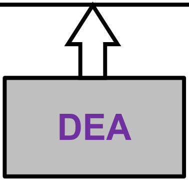
$$E_k^{in} = \frac{\sum_{m=1,\ldots,M} \mathbf{v_m} \cdot \mathbf{x_{mk}}}{\sum_{n=1,\ldots,N} \mathbf{u_n} \cdot \mathbf{y_{nk}}} = 1.097 \Rightarrow E_k^{out} = \frac{\sum_{n=1,\ldots,N} \mathbf{u_n} \cdot \mathbf{y_{nk}}}{\sum_{m=1,\ldots,M} \mathbf{v_m} \cdot \mathbf{x_{mk}}} = 0.911$$

efficiency is the reciprocal of optimum

Information Provided by DEA

- Objective measures of efficiency
- A reference set of comparable units
- Indicators of excess use of inputs
- Shortfalls in the production of outputs
- Returns to scale measure





Strong Points of DEA

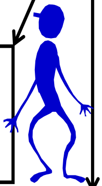
DEA

Performance evaluation and measurement

- No semantics: just deal with numbers
- No functional assumptions on inputs and outputs
- No assumption on the process: DMU is a black box

Benchmarking to identify "best practice" units

- Gives information about a DMU based on other DMUs
- "Maybe I can learn something from them"



"Data mining" to generate hypotheses about the drivers of efficiency

- Different units for inputs and outputs (no need for translation)
- No need to estimate weights of inputs and outputs (error-prone)

Weak Points of DEA

DEA

- Gives information about a DMU based on other DMUs (not against a theoretical maximum)

- Measures efficiencies relative to the efficient frontier that is defined by production possibilities
- This set may not be easy to characterize

- Poor discriminative power (many DMUs may be efficient)
- The more DMUs the more difficult it is to compare DMUs (DMUs become too specialized)
- Does not perform well with too many inputs/outputs (DMUs tend to be ”all optimal”)
- DEA scores reflect DMUs performance only for weights that are most favourable to it

Are Constant Returns to Scale Always Realistic?

A. Charnes, W. Cooper, E. Rhodes, Measuring the efficiency of decision making units, *European Journal of Operational Research*, 2(6), 429-444,1978

- So far, our formulation of DEA assumed Constant Returns to Scale (CRS)
- **Other formulations have been presented in order to drop this assumption**
- **Advantage**: most practical applications do not have CRS
- **Advantage**: besides telling us if a DMU is inefficient, they also tell us if it is too small or too large
- **Disadvantage**: efficiency scores change when we use input- vs. output-oriented versions (we have to choose an orientation for the model, according to the control we have over inputs vs. outputs)

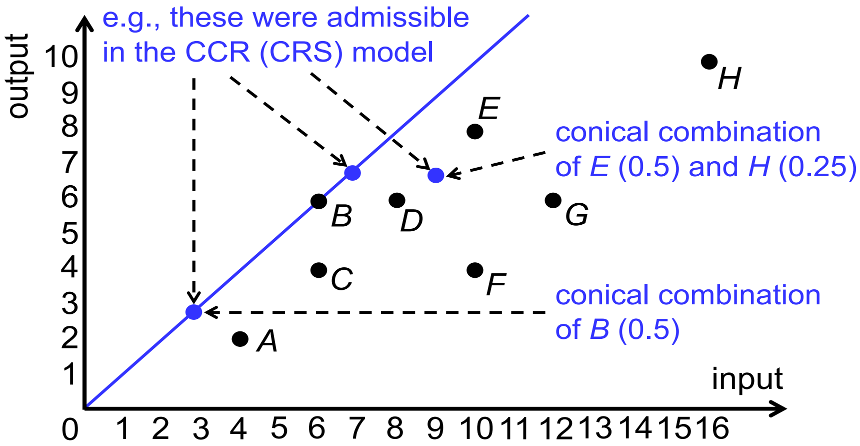
BCC Model – Variable Return to Scale (VRS)

R. Banker, A. Charnes, W. Cooper, Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis, *Management Science*, 30, 1078-1092,1984

- We consider economy of scale (scale of operations)
- We assume **Variable Returns to Scale** (**VRS**)
- **All convex combinations of existing DMUs are realistic**
- Each existing DMU can also either consume greater inputs and produce the same outputs or consume the same inputs and produce lesser outputs

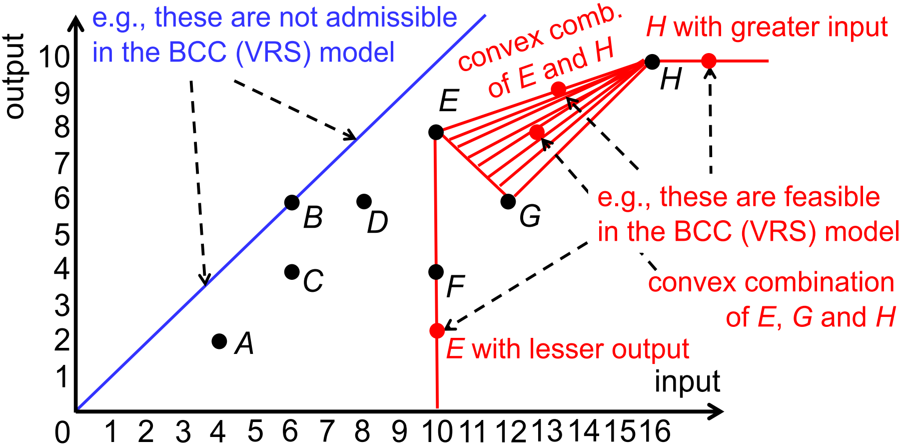
Change in the Producion Possibilities (1)

Store	DMU	A	B	C	D	E	F	G	H
Employees	**Input**	4	6	6	8	10	10	12	16
Cups of coffee	**Output**	2	6	4	6	8	4	6	10



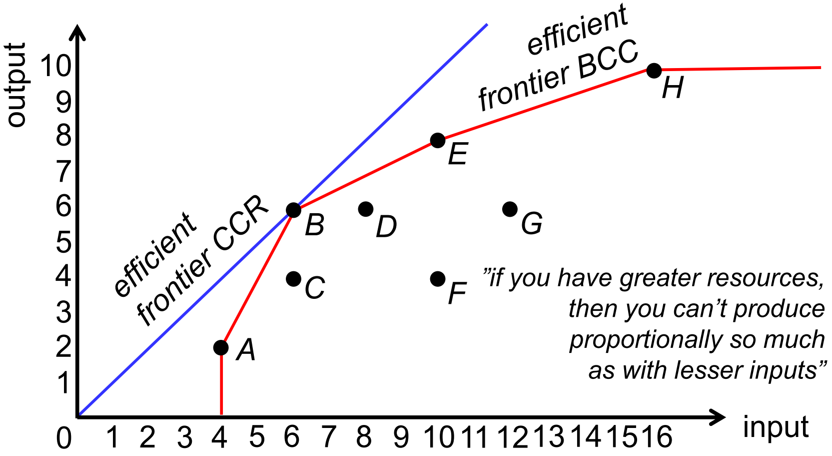
Change in the Producion Possibilities (2)

Store	DMU	A	B	C	D	E	F	G	H
Employees	**Input**	4	6	6	8	10	10	12	16
Cups of coffee	**Output**	2	6	4	6	8	4	6	10

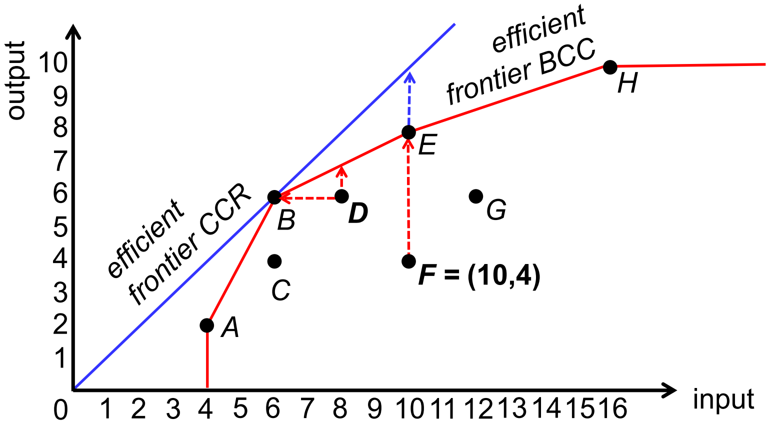


Efficient Frontiers in the BCC Model

Store	DMU	A	B	C	D	E	F	G	H
Employees	**Input**	4	6	6	8	10	10	12	16
Cups of coffee	**Output**	2	6	4	6	8	4	6	10



Comparison of BCC and CCR Models



- DMUs which are CCR-efficient are BCC-efficient
- DMUs which are BCC-efficient are not necessarily CCR-efficient
- Efficiency scores do differ (e.g., for F – $E_{\text{OUT-CCR}} = 4/10$; $E_{\text{OUT-BCC}} = 1/2$)
- BCC-efficiencies are orientation-dependant (e.g., for F – $E_{\text{IN-BCC}} = 6/8$; $E_{\text{OUT-BCC}} = 7/8$)

Input-oriented BCC models

$$\begin{aligned} & \min \theta_k \\ \text{s.t. } & \sum_{j=1,\dots,K} \lambda_j \cdot x_{mj} \le \theta_k \cdot x_{mk}, \quad m = 1, \dots, M \\ & \sum_{j=1,\dots,K} \lambda_j \cdot y_{nj} \ge y_{nk}, \quad n = 1, \dots, N \\ & \sum_{j=1,\dots,K} \lambda_j = 1 \\ & \theta_k \ge 0; \ \lambda_k \ge 0, \ k = 1, \dots, K \end{aligned}$$

$$\begin{aligned} & \max \sum_{n=1,\dots,N} u_n \cdot y_{nk} + u_0 \\ \text{s.t. } & \sum_{m=1,\dots,M} v_m \cdot x_{mk} = 1 \\ & \sum_{n=1,\dots,N} u_n \cdot y_{nj} + u_0 \le \sum_{m=1,\dots,M} v_m \cdot x_{mj} \\ & v_m \ge 0, \ m = 1, \dots, M \quad u_0 \text{ free (unrest.)} \\ & u_n \ge 0, \ n = 1, \dots, N \end{aligned}$$

Output-oriented BCC models

$$\begin{aligned} & \max \theta_k \\ \text{s.t. } & \sum_{j=1,\dots,K} \lambda_j \cdot x_{mj} \le x_{mk}, \quad m = 1, \dots, M \\ & \sum_{j=1,\dots,K} \lambda_j \cdot y_{nj} \ge \theta_k \cdot y_{nk}, \quad n = 1, \dots, N \\ & \sum_{j=1,\dots,K} \lambda_j = 1 \\ & \theta_k \ge 0; \ \lambda_k \ge 0, \ k = 1, \dots, K \end{aligned}$$

$$\begin{aligned} & \min \sum_{m=1,\dots,M} v_m \cdot x_{mk} + u_0 \\ \text{s.t. } & \sum_{n=1,\dots,N} u_n \cdot y_{nk} = 1 \\ & \sum_{m=1,\dots,M} v_m \cdot x_{mj} + u_0 \ge \sum_{n=1,\dots,N} u_n \cdot y_{nj} \\ & v_m \ge 0, \ m = 1, \dots, M \quad u_0 \text{ free (unrest.)} \\ & u_n \ge 0, \ n = 1, \dots, N \end{aligned}$$

BCC Efficiency Models – Example Formulations

$$\max\ 144 \cdot u_1 + 14400 \cdot u_2 + \boldsymbol{u_0}$$

$$\begin{aligned}
\text{s.t.}\quad & 8 \cdot v_1 = 1 \\
& 132 \cdot u_1 + 9600 \cdot u_2 + \boldsymbol{u_0} \le 6 \cdot v_1 \\
& 192 \cdot u_1 + 26400 \cdot u_2 + \boldsymbol{u_0} \le 12 \cdot v_1 \\
& 190 \cdot u_1 + 21000 \cdot u_2 + \boldsymbol{u_0} \le 10 \cdot v_1 \\
& 144 \cdot u_1 + 14400 \cdot u_2 + \boldsymbol{u_0} \le 8 \cdot v_1 \\
& v_1, u_1, u_2 \ge 0
\end{aligned}$$

DMU_k	In_1	Out_1	Out_2	**Eff.?**
DMU_1	**6**	**132**	**9600**	**Yes**
DMU_2	12	192	26400	**Yes**
DMU_3	10	190	21000	**Yes**
DMU_4	8	144	14400	**No**

$E_4 = 0.961$ (in CCR – 0.911)
DMU_4 is inefficient

$$\min\ \boldsymbol{\theta_4}$$

$$\begin{aligned}
\text{s.t.}\quad & 6 \cdot \lambda_1 + 12 \cdot \lambda_2 + 10 \cdot \lambda_3 + 8 \cdot \lambda_4 \le 8 \cdot \boldsymbol{\theta_4} \\
& 132 \cdot \lambda_1 + 192 \cdot \lambda_2 + 190 \cdot \lambda_3 + 144 \cdot \lambda_4 \ge 144 \\
& 9600 \cdot \lambda_1 + 26400 \cdot \lambda_2 + 21000 \cdot \lambda_3 + 14400 \cdot \lambda_4 \ge 14400 \\
& \boldsymbol{\theta_4} \ge 0;\ \lambda_k \ge 0,\ k = 1, \ldots, 4; \quad \boldsymbol{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1}
\end{aligned}$$

$\lambda_1 = 0.58$ $\lambda_3 = 0.42$
$\lambda_2 = \lambda_4 = 0$
$\theta_4 = 0.961$
$in_1 = 7.68 < 8$
$out_1 = 156.42 \ge 144$
$out_2 = 14400 \ge 14000$