Aalto University
School of Science

Data Envelopment Analysis (DEA) methods with focus on CCR-DEA

Mikko Rosenberg
Presentation *13*
5.11.2021

MS-E2191 Graduate Seminar on Operations Research
Fall 2021

Contents

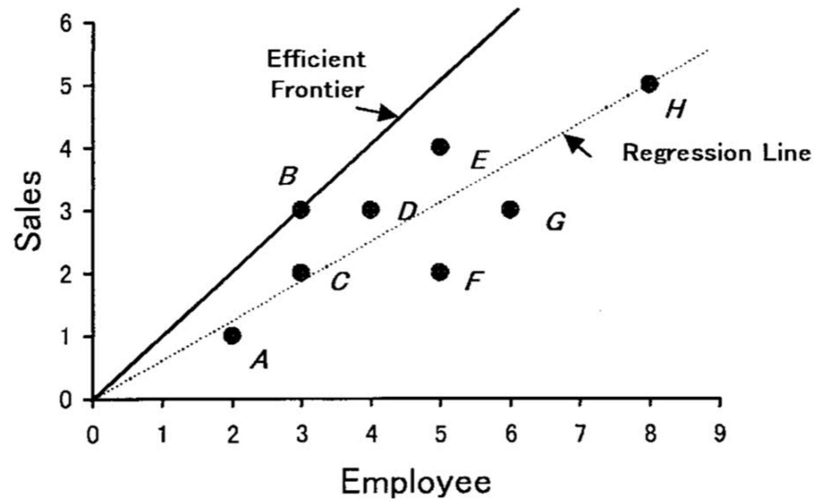
Data Envelopment Analysis (DEA)

- **Measure efficiency/productivity of decision making units DMUs**
  - Inputs: Number of employers, working hours, raw materials etc.
  - Outputs: monetary profit, products etc.
  - Efficiency: $\frac{\text{Output}}{\text{Input}}$
- **Input and output weights directly from data**
- **Mathematical programming**

Efficiency of a DMU

Store	A	B	C	D	E	F	G	H
Employee	2	3	3	4	5	5	6	8
Sale	1	3	2	3	4	2	3	5
Sale/Employee	0.5	1	0.667	0.75	0.8	0.4	0.5	0.625

Store	A	B	C	D	E	F	G	H
Efficiency	0.5	1	0.667	0.75	0.8	0.4	0.5	0.625



- **Efficient frontier: Reference to evaluate other DMUs**

$$\text{Efficiency of DMU j} = 0 \leq \frac{\text{Sales/Employee of j}}{\text{Sales/Employee of B}} \leq 1$$

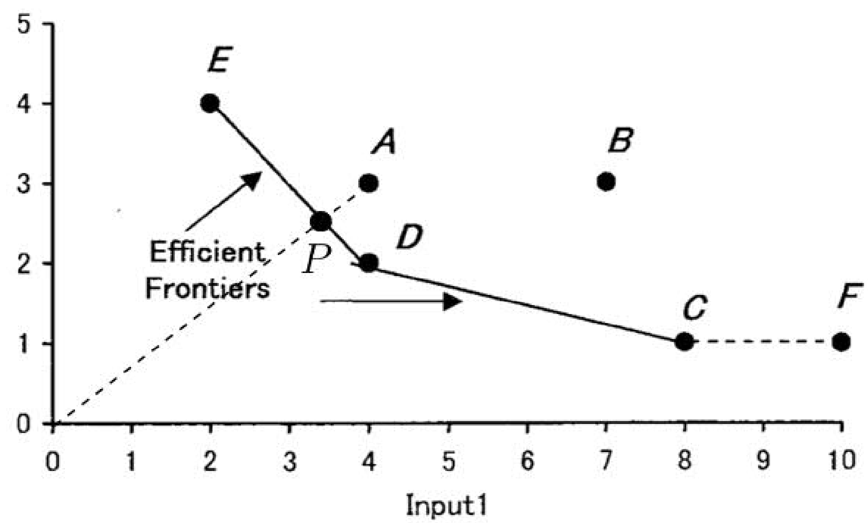
- **Units invariance**

Source: Introduction to Data Envelopment Analysis and Its Uses - With DEA-Solver Software and References, Cooper, W. W., Seiford, L. M., Tone, K, 2006

Example: Two inputs

- **Efficient frontier**
  - Connects all efficient points
  - Reference for efficiency
  - Efficiency of A: $\frac{OP}{OA}$
- **Production possibility set**
- **Reference set**
  - Determines the inefficiency of a given DMU
  - For A: {D,E}, For D: {D}

Store		A	B	C	D	E	F	G	H	I
Employee	x_1	4	7	8	4	2	5	6	5.5	6
Floor Area	x_2	3	3	1	2	4	2	4	2.5	2.5
Sale	y	1	1	1	1	1	1	1	1	1



Source: Introduction to Data Envelopment Analysis and Its Uses - With DEA-Solver Software and References, Cooper, W. W., Seiford, L. M., Tone, K, 2006

Inputs and outputs

- **Assume: $\boldsymbol{n}$ number of DMUs**
  - Denoted as DMU_j for $j = 1, \dots, n$
- **Each DMU has m inputs and s outputs**
- **Input data** $X \in \boldsymbol{R}_{>0}{}^{(m \times n)}$:
$$X = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{pmatrix}$$
- **Output data** $Y \in \boldsymbol{R}_{>0}{}^{(s \times n)}$:
$$Y = \begin{pmatrix} y_{11} & \cdots & y_{1n} \\ \vdots & \ddots & \vdots \\ y_{s1} & \cdots & y_{sn} \end{pmatrix}$$
- **Smaller inputs and higher outputs are preferred**

Partial and total efficiency

- **Benefits of total efficiency (simple fraction)**
- **How to obtain it?**

- **Non-negative weights for inputs $\boldsymbol{v} = (v_1, \dots, v_m)^T$ and outputs $\boldsymbol{u} = (u_1, \dots, u_s)^T$**
  - Virtual input $\sum_{i=1}^{m} v_i\, x_{io}$
  - Virtual output $\sum_{\mathrm{r}=1}^{s} u_r\, y_{ro}$

- **Efficiency:** $\frac{\text{Virtual output}}{\text{Virtual input}}$

CCR model – fractional problem (FP_o)

- **Idea: Find best possible weights u and v for each DMU relative to all other DMUs.**

- **Maximize the efficiency $\theta = \frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}}$ of DMU_o while assigning the same weights to all DMUs.**

Fractional problem for DMU_o:

$$\max_{v,u} \quad \theta = \frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}}$$

$$\text{s.t.:} \quad \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1 \; ; \quad j = 1, .., n$$

$$v_1, \ldots, v_m \geq 0$$

$$u_1, \ldots, u_s \geq 0$$

CCR model – linear problem (LP_o)

$$\max_{v,u} \quad \theta = \frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}}$$

$$\text{s.t.:} \quad \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1 \; ; \quad j = 1, .., n$$

$$v_1, ..., v_m \geq 0$$

$$u_1, ..., u_s \geq 0$$

→

$$\max_{u,v} \quad \theta = \sum_{r=1}^{s} u_r y_{ro}$$

$$\text{s.t.:} \quad \sum_{i=1}^{m} v_i x_{io} = 1$$

$$\sum_{r=1}^{s} u_r y_{yj} \leq \sum_{i=1}^{m} v_i x_{ij} \; ; \quad j = 1, .., n$$

$$v_1, ..., v_m \geq 0$$

$$u_1, ..., u_s \geq 0$$

- **Advantages of LP over FP:**
  - Can be solved with the simplex method
  - Existence of a dual problem

LP_o - CCR-efficiency

- **Solving $\mathbf{LP}_o$ gives the optimal values $\boldsymbol{\theta}^*$, $\boldsymbol{v}^*$ and $\boldsymbol{u}^*$.**
- **How to interpret these results?**

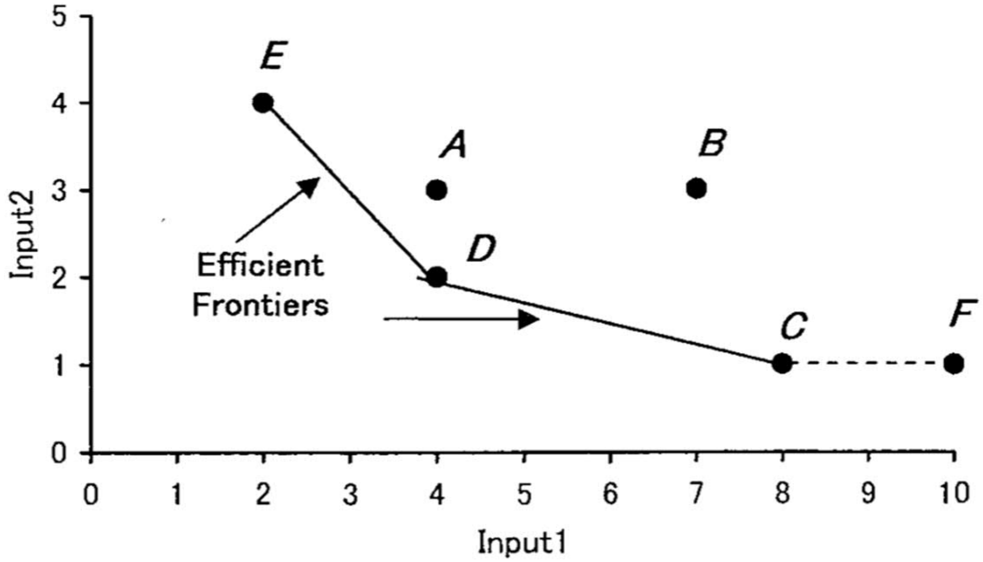
CCR-Efficiency

Definition: $\mathbf{DMU}_o$ is CCR-efficient if both

- $\theta^* = 1$
- Exists at least one optimal $(\boldsymbol{v}^*, \boldsymbol{u}^*)$ for which $\boldsymbol{v}^* > \mathbf{0}$ and $\boldsymbol{u}^* > \mathbf{0}$

- **What if $\boldsymbol{v}^* = \mathbf{0}$ and/or $\boldsymbol{u}^* = \mathbf{0}$?**

DMU	x_1	x_2	y	CCR(θ^*)	Reference Set	v_1	v_2	u
A	4	3	1	0.8571	D E	.1429	.1429	.8571
B	7	3	1	0.6316	C D	.0526	.2105	.6316
C	8	1	1	1	C	.0833	.3333	1
D	4	2	1	1	D	.1667	.1667	1
E	2	4	1	1	E	.2143	.1429	1
F	10	1	1	1	C	0	1	1



Source: Introduction to Data Envelopment Analysis and Its Uses - With DEA-Solver Software and References, Cooper, W. W., Seiford, L. M., Tone, K, 2006

Dual problem (DLP$_o$) of LP$_o$

$$(\mathrm{LP}_o): \quad \max_{\boldsymbol{v},\boldsymbol{u}} \; \boldsymbol{u}^\intercal \boldsymbol{y}_o$$
$$\text{s.t.:} \quad \boldsymbol{v}^\intercal \boldsymbol{x}_o = 1$$
$$-\boldsymbol{v}^\intercal X + \boldsymbol{u}^\intercal Y \leq \mathbf{0}$$
$$\boldsymbol{v} \geq \mathbf{0}$$
$$\boldsymbol{u} \geq \mathbf{0}$$

$$(\mathrm{DLP}_o): \quad \min_{\theta \in \mathbb{R}, \boldsymbol{\lambda} \in \mathbb{R}^n} \; \theta$$
$$\text{s.t.:} \quad \theta \boldsymbol{x}_o - X\boldsymbol{\lambda} \geq \mathbf{0}$$
$$Y\boldsymbol{\lambda} \geq \boldsymbol{y}_o$$
$$\boldsymbol{\lambda} \geq \mathbf{0}$$

Strong duality theorem: Optimal objective value is same for the primal and the dual problem.

Input excesses and output shortfalls

- Excesses of the inputs $\boldsymbol{s}^{-} \in \boldsymbol{R}^{m}$ and shortfalls of the outputs $\boldsymbol{s}^{+} \in \boldsymbol{R}^{s}$ are defined as

$$\boldsymbol{s}^{-} = \theta \boldsymbol{x}_o - X\boldsymbol{\lambda}$$
$$\boldsymbol{s}^{+} = Y\boldsymbol{\lambda} - \boldsymbol{y}_o$$

- Slack vectors of the inequalities in the $\mathbf{DLP}_{\boldsymbol{o}}$

$$\theta \boldsymbol{x}_o - X\boldsymbol{\lambda} \geq \mathbf{0}$$
$$Y\boldsymbol{\lambda} \geq \boldsymbol{y}_o$$

- Must be non-negative for a feasible solution

Two phase LP problem

- **Phase I: Solve $\mathbf{DLP}_{\boldsymbol{o}}$ and use θ^* in next phase**
- **Phase II: maximize input excesses and output shortfalls**

  - $\boldsymbol{e}^{\boldsymbol{T}}$ are vectors of ones
  - Solution gives $\boldsymbol{\lambda}^*, \boldsymbol{s}^-$ and $\boldsymbol{s}^+$
  - Zero-slack if $\boldsymbol{s}^- = \boldsymbol{s}^+ = \mathbf{0}$

$$
\begin{aligned}
\max_{\boldsymbol{\lambda}, \boldsymbol{s}^-, \boldsymbol{s}^+} \quad & \omega = \boldsymbol{e}^\intercal \boldsymbol{s}^- + \boldsymbol{e}^\intercal \boldsymbol{s}^+ \\
\text{s.t.:} \quad & \boldsymbol{s}^- = \theta^* \boldsymbol{x}_o - X\boldsymbol{\lambda} \\
& \boldsymbol{s}^+ = Y\boldsymbol{\lambda} - \boldsymbol{y}_o \\
& \boldsymbol{\lambda}, \boldsymbol{s}^-, \boldsymbol{s}^+ \geq \mathbf{0}
\end{aligned}
$$

CCR-efficiency revisited

Previously:

- $\mathbf{DMU}_{\boldsymbol{o}}$ **is CCR-efficient if both**
  - $\theta^* = 1$
  - Exists at least one optimal $(\boldsymbol{v}^*, \boldsymbol{u}^*)$ for which $\boldsymbol{v}^* > \mathbf{0}$ and $\boldsymbol{u}^* > \mathbf{0}$

Now:

- $\mathbf{DMU}_{\boldsymbol{o}}$ **is CCR-efficient if both**
  - $\theta^* = 1$
  - All slacks are zero (zero-slack solution)

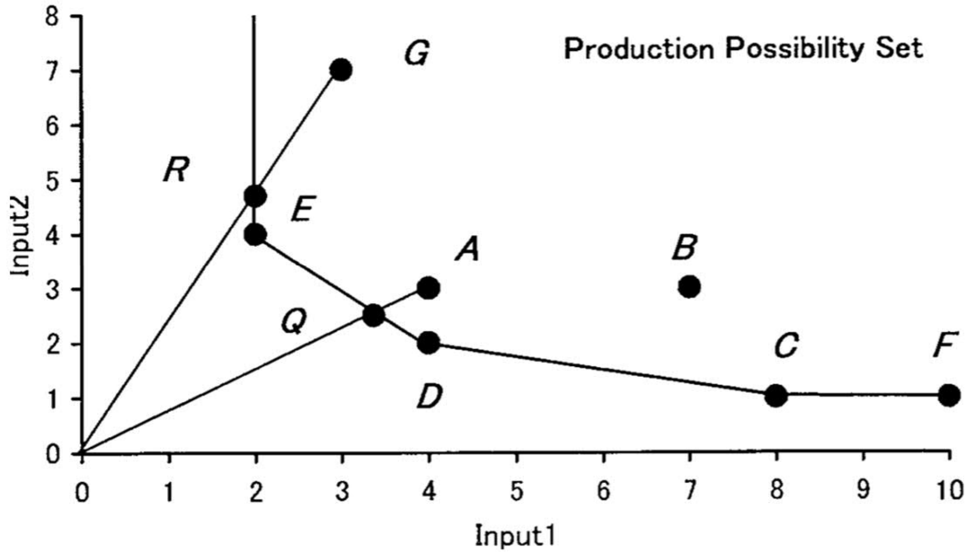
Both definitions are equivalent

Reference sets

- **Solving phase II gives $\boldsymbol{\lambda}^*$**
- **Reference set of $\mathbf{DMU}_{\boldsymbol{o}}$ is given by**

$$E_o = \{j | \boldsymbol{\lambda}_j^* > 0\} \quad (j = 1, ..., n)$$

DMU	CCR-Eff θ^*	Ref Set	Excess s_1^-	s_2^-	Shortfall s^+
A	0.8571	D E	0	0	0
B	0.6316	C D	0	0	0
C	1.0000	C	0	0	0
D	1.0000	D	0	0	0
E	1.0000	E	0	0	0
F	1.0000	C	2	0	0
G	0.6667	E	0	.6667	0



A: λ^* = [0, 0, 0, 0.71, 0.29, 0, 0]

C: λ^* = [0, 0, 1, 0, 0, 0, 0]

F: λ^* = [0, 0, 1, 0, 0, 0, 0]

Source: Introduction to Data Envelopment Analysis and Its Uses - With DEA-Solver Software and References, Cooper, W. W., Seiford, L. M., Tone, K, 2006

Advantages of using $\mathbf{DLP}_{\boldsymbol{o}}$

- **Linear programming problem grows in powers of number of constraints**
- $\mathbf{LP}_{\boldsymbol{o}}$ **has** n **constraints**
- $\mathbf{DLP}_{\boldsymbol{o}}$ **has** $(m + s)$ **constraints**
- $\mathbf{DLP}_{\boldsymbol{o}}$ **is computationally better if the number of DMUs is larger than sum of inputs and outputsand**
- **Solution interpretation**

References

1. Cooper, W. W., Seiford, L. M., Tone, K., 2006: Introduction to Data Envelopment Analysis and Its Uses - With DEA-Solver Software and References, Springer Science+Business Media, New York.
2. Charnes, A., Cooper, W. W., Rhodes, E., 1978: Measuring the efficiency of decision making units, European Journal of Operational Research 2/6, s. 429-444.

Homework (1/3)

Consider the following six DMUs with two inputs x_1 and x_2 and an output y. The optimal values θ^*, λ^*, s^- and s^+ are solved using the two phase linear problem (slide 14).

DMU	x_1	x_2	y	θ^*	λ^*	$s^-{}_1$	$s^-{}_2$	s^+
A	5	3	1	1	[1 0 0 0 0 0]	0	0	0
B	3	8	1	1	[0, 0, 0, 1, 0, 0]	0	3	0
C	6	6	1	0.67	[0.5, 0, 0, 0.5, 0, 0]	0	0	0
D	3	5	1	1	[0, 0, 0, 1, 0, 0]	0	0	0
E	9	2	1	1	[0, 0, 0, 0, 1, 0]	0	0	0
F	9	5	1	0.59	[0.93, 0, 0, 0, 0.06, 0]	0	0	0

Homework (2/3)

1. **Plot the DMUs using the inputs (see slide 17) and answer the following questions**
   a) Which DMUs are efficient and which inefficient? What is the efficient frontier of the problem (plot/describe)?
   b) What are the reference sets for DMUs A, C and F?
2. **Based on the values of θ^*, λ^*, s^- and s^+ answer the following questions. Justify your answers.**
   a) Which DMUs are CCR-efficient and which CCR-inefficient?
   b) What are the reference sets for DMUs A, B and C?

Homework (3/3)

3. **How would you change the inputs x_1 and x_2 to**
   a) make the alphabetically first ineficcient DMU efficient?
   b) make the alphabetically first CCR-inefficient DMU CCR-efficient?

Email answers to mikko.rosenberg@aalto.fi by 12.11. 9am.

Use preferably subject " MS-E2191 HW13".

You can ask any questions regarding the homework or materials via email or tg @rosenlew.