Data Envelopment Analysis

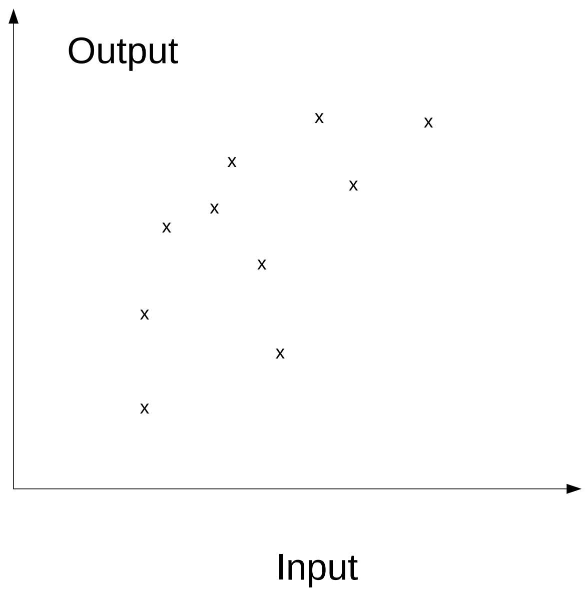
Georg Pflug
University of Vienna

EUMOptFin 3, Bergamo

The problem

Input data points: $x^{(1)}, \ldots, x^{(n)} \in \mathbb{R}^m$
Output data points: $y^{(1)}, \ldots, y^{(n)} \in \mathbb{R}$

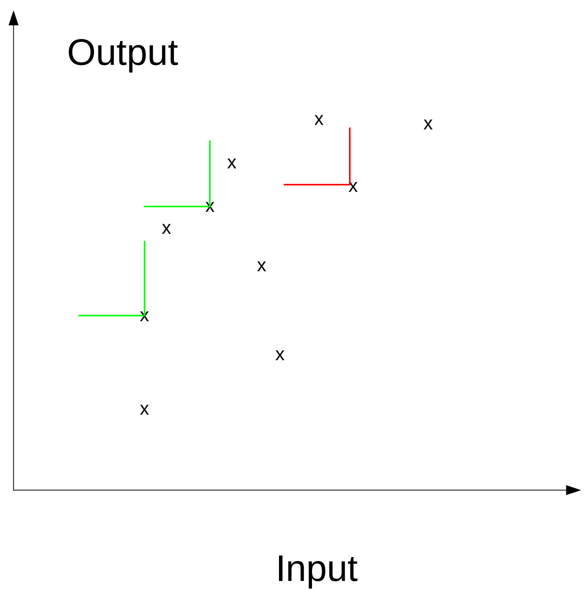
representing the operational characteristics of n decision making units (DMU's).



Definition. A DMU (i) is called non-dominated, if there is no other DMU (j) such that

$$\begin{array}{ccc} x_1^{(j)} & \leq & x_1^{(i)} \\ x_2^{(j)} & \leq & x_2^{(i)} \\ & \vdots & \\ x_m^{(j)} & \leq & x_m^{(j)} \\ y^{(j)} & \geq & y^{(j)} \end{array}$$

with at least one inequality sign.



Production functions

Let $\mathcal{F} = \{f(x, \nu) : \nu \in M\}$ be a family of input-output functions $\mathbb{R}^m \to \mathbb{R}$.

Examples are:

(a) linear function

$$f(x, \nu) = \sum_{i=1}^{m} \nu_i x_i$$

(b) Affine-linear function

$$f(x, \nu) = \nu_0 + \sum_{i=1}^{m} \nu_i x_i$$

(c) Cobb-Douglas function

$$f(x, \nu) = \prod_{i=1}^{m} x_i^{\nu_i}$$

(d) Constant elasticity of substitution (CES - function)

$$f(x,\nu) = \left(\sum_{i=1}^{m} \nu_i x_i^{-\rho}\right)^{-1/\rho}$$

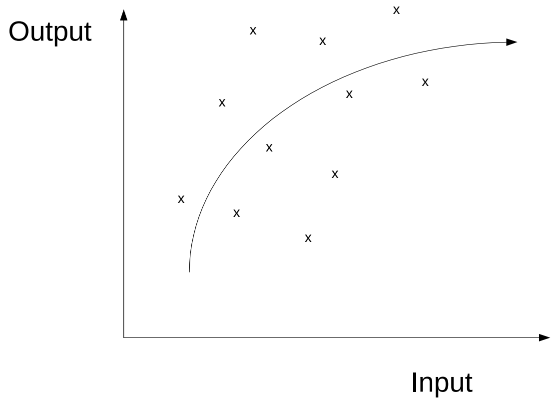
(e) Loglinear function

$$f(x,\nu) = \sum_{i=1}^{m} \nu_i \log(x_i)$$

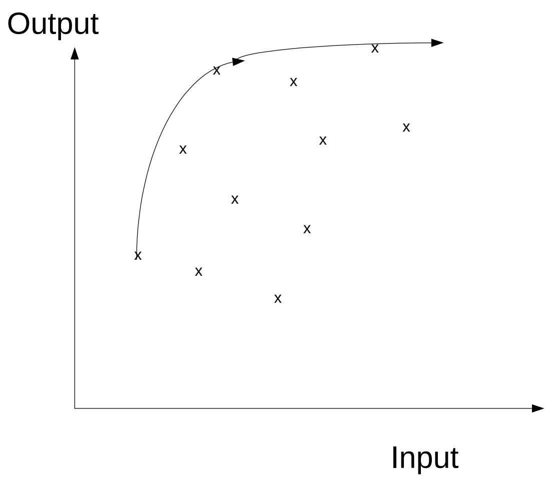
(f) Some nonlinear production function

$$f(x,\nu) = \nu_1 x_1 (\nu_2 x_2 + x_3).$$

Regression functions and envelope functions



A regression function



An envelope function

$X = (x^{(1)}, \ldots, x^{(n)})$

$y = (y^{(1)}, \ldots, y^{(n)})$

Let $\mathcal{F} = \{f(x, \nu) : \nu \in M\}$ be a family of input-output functions. The *envelope function* of the data set (X, y) is defined as

$$\bar{f}(x) = \min\{f(x, \nu) : \nu \text{ is such that for all } j \ \ f(x^{(j)}, \nu) \geq y^{(j)}\}$$

In order to find whether a pair $x^{(i)}, y^{(i)}$ lies on the envelope functions, one solves

$$\min\{f(x^{(i)}, \nu) : \nu \text{ is such that for all } j \ \ f(x^{(j)}, \nu) \geq y^{(j)}\}$$

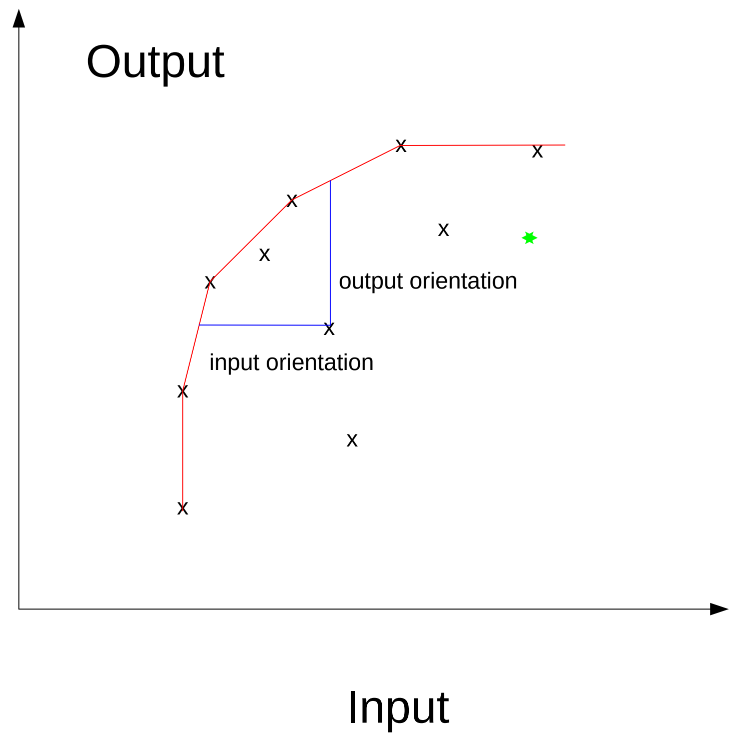
and compares this with $y^{(i)}$.

The *input oriented degree of efficiency* of unit i is

$$\min\{\theta : \bar{f}(\theta x^{(i)}) \geq y^{(i)}\}.$$

The *output oriented degree of efficiency* of unit i is

$$\max\{\phi : \bar{f}(x^{(i)}) \geq \phi y^{(i)}\} = \bar{f}(x^{(i)})/y^{(i)}.$$


Output
output orientation
input orientation
Input

Multiple output

If there are more than one outputs, then one uses an aggregation function $y^{(i)} = \sum_{j=1}^{s} \mu_j y_j^{(i)}$.

The *input oriented degree of efficiency* of unit i is

$$\min\{\theta(\mu_1, \dots, \mu_s) : \theta(\mu_1, \dots, \mu_s) \text{ is the efficiency for the aggregation function } \sum_{j=1}^{s} \mu_j y_j^{(i)}\}.$$

i.e. DMU i is input oriented efficient for multiple output, if it is input oriented efficient for all aggregation functions.

The *output oriented degree of efficiency* of unit i is

$$\max\{\phi(\mu_1, \dots, \mu_s) : \phi(\mu_1, \dots, \mu_s) \text{ is the efficiency for the aggregation function } \sum_{j=1}^{s} \mu_j y_j^{(i)}\}.$$

i.e. DMU i is output oriented efficient for multiple output, if there is at least one aggregation function such that it is output oriented efficient.

Linear Models

We form the $m \times n$ matrix $X = (x_i^{(j)})$ and the $s \times n$ matrix $Y = (y_i^{(j)})$. The i-st column, corresponding to the i-st DMU of X and Y is denoted by X_i resp. Y_i.

Duality of linear programs:

$$\begin{aligned} \max\ & c_1^T x + c_2^T y \\ & A_{11}x + A_{12}y \leq b_1 \\ & A_{21}x + A_{22}y = b_2 \\ & x \geq 0;\ y \text{ arbitrary} \end{aligned}$$

$$\begin{aligned} \min\ & b_1^T u + b_2^T v \\ & A_{11}^T u + A_{21}^T v \geq c_1 \\ & A_{12}^T u + A_{22}^T v = c_2 \\ & u \geq 0;\ v \text{ arbitrary} \end{aligned}$$

Specific models

Affine linear model (variable returns to scale - VRS)

The affine linear family produces the convex hull as envelope function.

$$f(x, \nu) = \nu^T x + u_0; \nu \geq 0.$$

Input orientation:

$$\begin{aligned} \min \theta \\ \theta \nu^T X_i + u_0 - \mu^T Y_i \geq 0 \\ \mu^T Y - \nu^T X - u_0 \mathbb{1} \leq 0 \\ \mu \geq 0 \\ \nu \geq 0 \end{aligned}$$

which is equivalent to

$$\begin{aligned} \max \mu^T Y_i - u_0 \\ \nu^T X_i = 1 \\ \mu^T Y - \nu^T X - u_0 \mathbb{1} \leq 0 \\ \mu \geq 0 \\ \nu \geq 0 \end{aligned}$$

The dual is

$$
\begin{aligned}
\min \theta \\
Y\lambda \geq Y_i \\
\theta X_i \geq \lambda X \\
\mathbb{1}\lambda = 1 \\
\lambda \geq 0
\end{aligned}
$$

Once the optimal θ has been determined, we solve the following problem to identify the slacks:

$$
\begin{aligned}
\max \mathbb{1}^T s^+ + \mathbb{1}^T s^- \\
Y\lambda - s^+ - Y_i = 0 \\
\theta X_i - X\lambda - s^- = 0 \\
\mathbb{1}\lambda = 1 \\
\lambda, s^+, s^- \geq 0
\end{aligned}
$$

Output orientation:

$$\begin{aligned}
\max \phi \\
\nu^T X_i + u_0 \geq \phi(\mu^T Y_i) \\
\mu^T Y - \nu^T X - u_0 \mathbb{1} \leq 0 \\
\mu \geq 0 \\
\nu \geq 0
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
\min \nu^T X_i + u_0 \\
\mu^T Y_i = 1 \\
\mu^T Y - \nu^T X - u_0 \mathbb{1} \leq 0 \\
\mu \geq 0 \\
\nu \geq 0
\end{aligned}$$

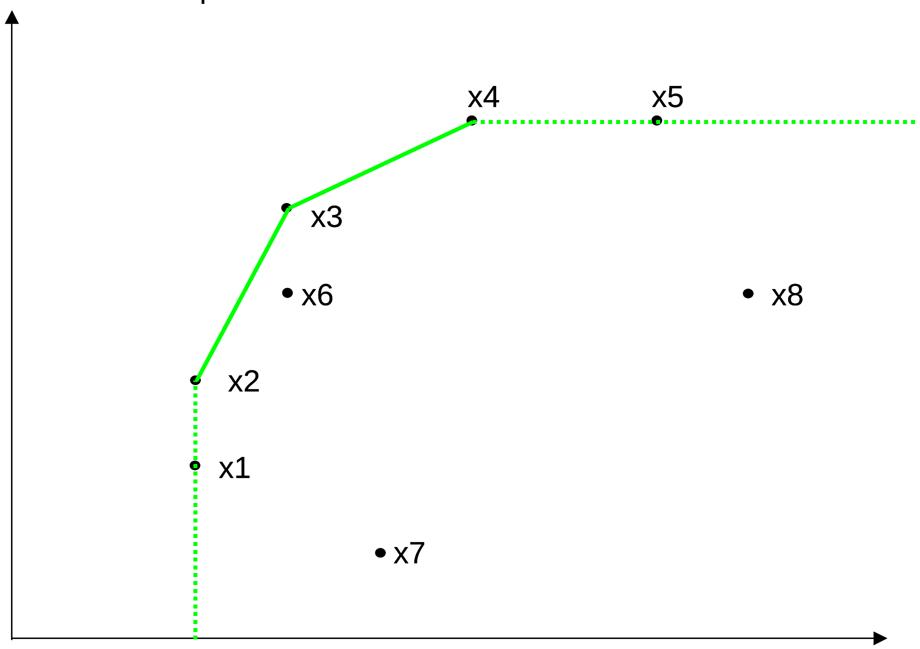
The dual is

$$\begin{aligned}
\max \phi \\
\phi Y_i \leq \lambda Y \\
X\lambda \leq X_i \\
\mathbb{1}\lambda = 1 \\
\lambda \geq 0
\end{aligned}$$

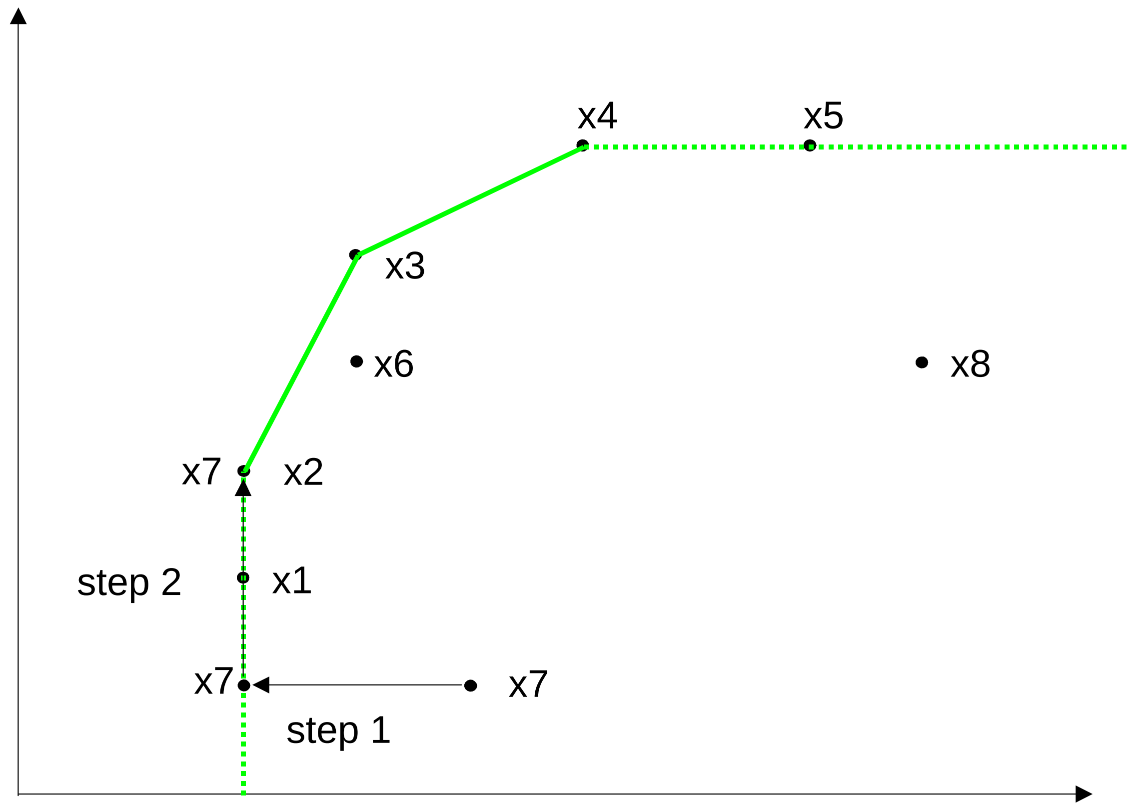
Once the optimal ϕ has been determined, we solve

$$\begin{aligned} \max \mathbf{1}^T s^+ + \mathbf{1}^T s^- \\ \theta Y_i - Y\lambda + s^+ = 0 \\ X\lambda + s- = X_i \\ \mathbf{1}^T \lambda = 1 \\ \lambda, s^+, s^- \geq 0 \end{aligned}$$

BCCpi



	X	Y
P1	2	2
P2	2	3
P3	3	5
P4	5	6
P5	7	6
P6	3	4
P7	4	1
P8	8	4


x4
x5
x3
x6
x8
x7
x2
step 2
x1
x7
x7
step 1

Linear model (constant returns to scale - CRS):

$$f(x, \nu) = \nu^T x; \nu \geq 0.$$

The linear family produces the conical hull as envelope function.

Input orientation:

$$\begin{aligned} \min \theta \\ \theta \nu^T X_i - \mu^T Y_i \geq 0 \\ \mu^T Y - \nu^T X \leq 0 \\ \mu \geq 0 \\ \nu \geq 0 \end{aligned}$$

which is equivalent to

$$\begin{aligned} \max \mu^T Y_i \\ \nu^T X_i = 1 \\ \mu^T Y - \nu^T X \mathbf{1} \leq 0 \\ \mu \geq 0 \\ \nu \geq 0 \end{aligned}$$

The dual is

$$\begin{aligned} \min \theta \\ Y\lambda - Y_i \geq 0 \\ \theta X_i - \lambda X \geq 0 \\ \lambda \geq 0 \end{aligned}$$

Once the optimal θ has been determined, we solve the following problem to identify the slacks:

$$\begin{aligned} \max \mathbb{1}^T s^+ + \mathbb{1}^T s^- \\ Y\lambda - s^+ - Y_i = 0 \\ \theta X_i - \lambda X - s- = 0 \\ \lambda, s^+, s^- \geq 0 \end{aligned}$$

Output orientation:

$$\begin{aligned} \max \phi \\ \nu^T X_i \geq \phi(\mu^T Y_i) \\ \mu^T Y - \nu^T X\mathbf{1} \leq 0 \\ \mu \geq 0 \\ \nu \geq 0 \end{aligned}$$

which is equivalent to

$$\begin{aligned} \min \nu^T X_i \\ \mu^T Y_i = 1 \\ -\mu^T Y + \nu^T X \geq 0 \\ \mu \geq 0 \\ \nu \geq 0 \end{aligned}$$

The dual is

$$\begin{aligned} \max \phi \\ -\phi Y_i + \lambda Y \geq 0 \\ -X\lambda + X_i \geq 0 \\ \lambda \geq 0 \end{aligned}$$

Once the optimal ϕ has been determined, we solve

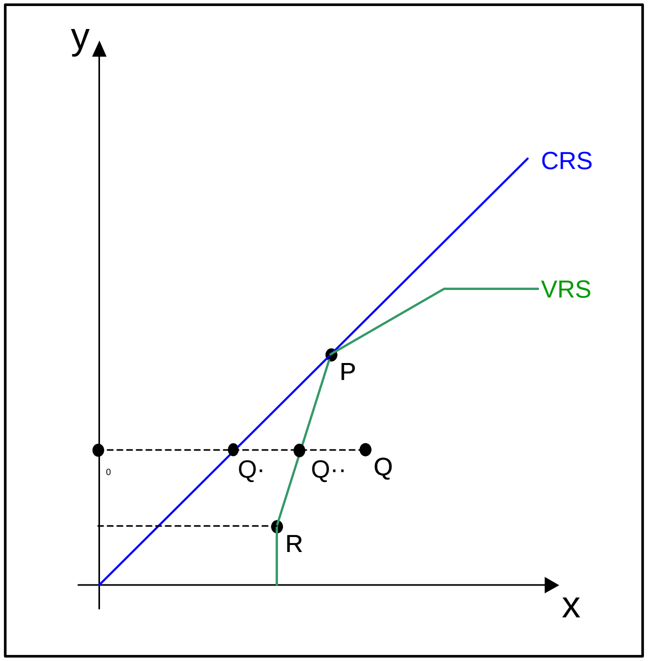
$$
\begin{aligned}
\max \mathbf{1}^T s^+ + \mathbf{1}^T s^- \\
\theta Y_i - Y\lambda + s^+ = 0 \\
X\lambda + s- = X_0 \\
\lambda, s^+, s^- \geq 0
\end{aligned}
$$

Special notions of efficiency

pure technical efficiency $PTE(i)$: Efficiency in the input oriented VRS model.

technical efficiency $TE(i)$: Efficiency in the input oriented CRS model.

Scale efficiency: $SE(i) = TE(i)/PTE(i) \leq 1$, since $PTE(i) \geq TE(i)$.



$$SE \quad \frac{TE}{PTE}$$

$$TE_Q \quad 0Q^{\cdot}/0Q$$

$$PTE_Q \quad 0Q^{\cdot\cdot}/0Q$$

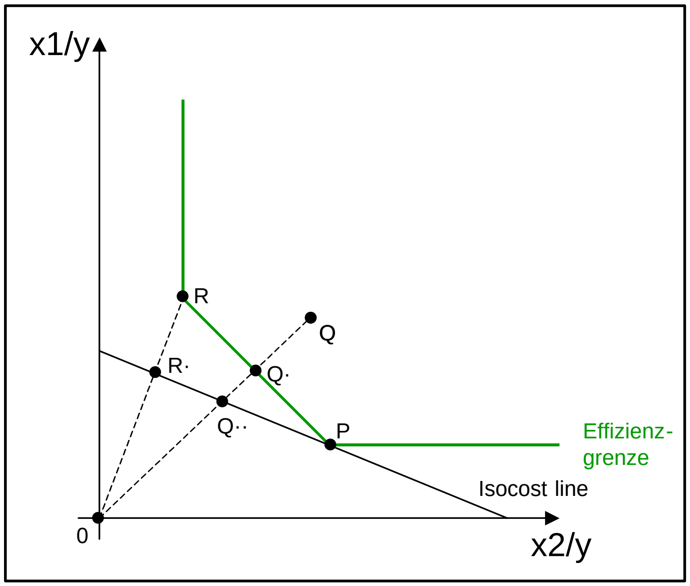
$$SE_Q \quad 0Q^{\cdot}/0Q^{\cdot\cdot}$$

DMU	TE	PTE	SE
P	1	1	1
Q	0,5	0,7	0,71
R	0,3	1	0,3

$$TE = 0Q^{\cdot}/0Q \;\; PTE = 0Q^{\cdot\cdot}/0Q \;\; SE = 0Q^{\cdot}/0Q^{\cdot\cdot}$$

Cost efficiency: $CE(i) = [\min \sum_j x_j p_j : \bar{f}(x) \geq y^{(i)}] / [\sum_j x_j^{(i)} p_j]$

Allocation efficiency: $AE(i) = CE(i)/TE(i)$



$AE \quad \frac{CE}{TE}$

$TE_Q \quad 0Q\cdot/0Q$

$CE_Q \quad 0Q\cdot\cdot/0Q$

$AE_Q \quad 0Q\cdot/0Q\cdot\cdot$

DMU	TE	CE	AE
P	1	1	1
Q	0,7	0,5	0,714
R	1	0,75	0,75

$$TE = 0Q^{\cdot}/0Q \; CE = 0Q^{\cdot\cdot}/0Q \; AE = 0Q^{\cdot}/0Q^{\cdot\cdot}$$