Data Envelopment Analysis (Analýza obalu dat)

Martin Branda

Univerzita Karlova v Praze
Matematicko-fyzikální fakulta
Katedra pravděpodobnosti a matematické statistiky

Optimalizace s aplikací ve financích

Motivace

3 firmy:

	1	2	3
Suroviny (vstupy)	3	6	3
Výrobky (výstupy)	7	11	5

Která pracuje „nejlépe“ – efektivně – eficientně?

$$\frac{7}{3} > \frac{11}{6} > \frac{5}{3},$$

tedy (asi) firma 1.

- Co když je vstupů a výstupů více?
- Co když zdvojnásobením vstupů nemůžu zdvojnásobit výrobu?

Decision Making Units

- **Homogenní jednotky** – Decision Making Units (DMU) $j = 1, \ldots, n$
- **Vstupy** $X = \{x_{ij}\}$, $i = 1, \ldots, m$ (preferujeme nižší hodnoty)
- **Výstupy** $Y = \{y_{rj}\}$, $r = 1, \ldots, s$ (preferujeme vyšší hodnoty)

Předpokládáme, že data jsou kladná.

Příklad – bankovní pobočky

Vstupy

- počet zaměstnanců (dále děleno dle kvalifikace: junior, senior, vedoucí)
- rozloha pobočky
- nemzdové náklady

Výstupy

- počet uzavřených smluv (běžný účet, hypotéka, spotřebitelská půjčka, pojištění)
- počet nově získaných klientů

Příklad – fakulty

Vstupy

- počet zaměstnanců (dále děleno dle kvalifikace: asistent, docent, profesor)
- počet studentů, kteří nastoupí do 1. ročníku

Výstupy

- počet vědeckých publikací
- počet absolventů (Bc., Mgr., Ph.D.)

Posouzení eficience jednotky $0 \in \{1, \dots, n\}$

$$\max_{u_r, v_i} \frac{\sum_{r=1}^{s} u_r y_{r0}}{\sum_{i=1}^{m} v_i x_{i0}}$$

$$\begin{aligned} &\text{s.t.} \\ \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} &\leq 1,\ j = 1, \dots, n, \\ u_r &\geq 0,\ r = 1, \dots, s, \\ v_i &\geq 0,\ i = 1, \dots, m. \end{aligned}$$

Jednotka 0 je eficientní, právě když je optimální hodnota rovna jedné.
Každá jednotka dostane pro ni nejvýhodnější váhy.

Položíme

$$
\begin{aligned}
t &= \frac{1}{\sum_{i=1}^{m} v_i x_{i0}}, \\
\tilde{u}_r &= t \cdot u_r, \\
\tilde{v}_i &= t \cdot v_i,
\end{aligned}
$$

čímž odstraníme podíly a získáme úlohu LP ..

Po Charnesově–Cooperově transformaci LP:

$$
\begin{aligned}
\max_{u_r, v_i} \sum_{r=1}^{s} u_r y_{r0} & & \\
& \text{s.t.} & \\
\sum_{i=1}^{m} v_i x_{i0} &= 1, \\
\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} &\le 0, \; j = 1, \dots, n, \\
u_r &\ge 0, \; r = 1, \dots, s, \\
v_i &\ge 0, \; i = 1, \dots, m.
\end{aligned}
$$

Dualita v LP (na cvičení) ...

$$
\begin{aligned}
&\min_{\theta,\lambda_j} \theta \\
&\qquad \text{s.t.} \\
\sum_{j=1}^{n} \lambda_j x_{ij} \quad &\leq \quad \theta x_{i0},\ i = 1, \ldots, m, \\
\sum_{j=1}^{n} \lambda_j y_{rj} \quad &\geq \quad y_{r0},\ r = 1, \ldots, s, \\
\lambda_j \quad &\geq \quad 0,\ j = 1, \ldots, n.
\end{aligned}
$$

Model orientovaný na vstupy.

Infinitezimální $\varepsilon > 0$, aby byly všechny vstupy a výstupy zahrnuty

$$
\begin{aligned}
\max_{u_r, v_i} \sum_{r=1}^{s} u_r y_{r0} & & \\
& \text{s.t.} & \\
\sum_{i=1}^{m} v_i x_{i0} &= 1, \\
\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} &\leq 0,\ j = 1, \dots, n, \\
u_r &\geq \varepsilon,\ r = 1, \dots, s, \\
v_i &\geq \varepsilon,\ i = 1, \dots, m.
\end{aligned}
$$

Charnes–Cooper–Rhodes (CCR) model

Duální (obalová) formulace

$$
\begin{aligned}
\min_{\theta,\lambda_j,s_i^-,s_r^+} \theta \quad & - \quad \varepsilon \left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ \right) \\
& \text{s.t.} \\
\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- \quad & = \quad \theta x_{i0}, \ \ i = 1, \ldots, m, \\
\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ \quad & = \quad y_{r0}, \ \ r = 1, \ldots, s, \\
\lambda_j \quad & \geq \quad 0, \ j = 1, \ldots, n, \\
s_i^- \quad & \geq \quad 0, \ i = 1, \ldots, m, \\
s_r^+ \quad & \geq \quad 0, \ r = 1, \ldots, s.
\end{aligned}
$$

Klasifikace jednotek

Nechť θ^*, s_i^{-*}, s_r^{+*} je optimální řešení, potom jednotka je

- Neeficientní $\theta^* < 1$.
- Slabě eficientní $\theta^* = 1$ a existuje $s_i^{-*} > 0$ nebo $s_r^{+*} > 0$.
- Silně eficientní $\theta^* = 1$ a všechny $s_i^{-*} = 0$ nebo $s_r^{+*} = 0$.

Production possibility set

Množina možných produktů

$$\mathcal{PPS} = \left\{ (x, y) : \; x_i = \sum_{j=1}^{n} \lambda_j x_{ij}, \; y_r = \sum_{j=1}^{n} \lambda_j y_{rj}, \; \lambda_j \geq 0 \right\}$$

Rostou se vstupy proporcionálně i výstupy? Tj.

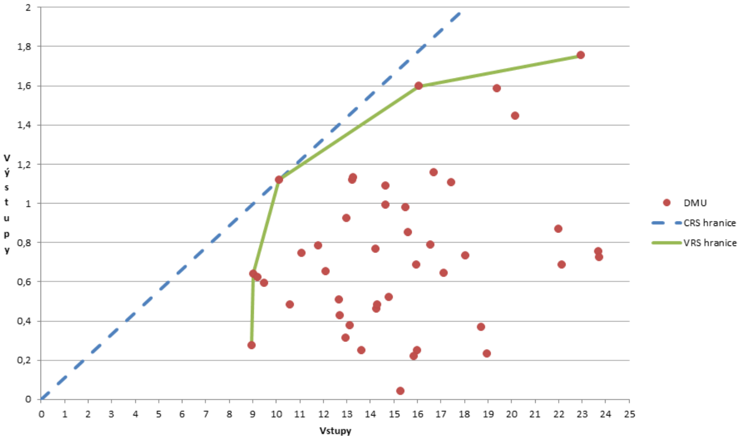
- $(x, y) \in \mathcal{PPS},\ \alpha > 0 \Longrightarrow (\alpha x, \alpha y) \in \mathcal{PPS}$?

Platí-li, **konstantní výnosy z rozsahu** (Constant Returns to Scale – CRS).

Neplatí-li, **variabilní výnosy z rozsahu** (Variable Returns to Scale – VRS):

$$\mathcal{PPS}^{VRS} = \left\{ (x, y) :\ x_i = \sum_{j=1}^{n} \lambda_j x_{ij},\ y_r = \sum_{j=1}^{n} \lambda_j y_{rj},\ \sum_{\mathbf{j=1}}^{\mathbf{n}} \boldsymbol{\lambda}_{\mathbf{j}} = 1,\ \lambda_j \geq 0 \right\}$$

Výnosy z rozsahu



Výnosy z rozsahu

Z obrázku vidíme:

- CRS – nejmenší konvexní kužel obsahující data
- VRS – „horní" konvexní obal dat

Banker–Charnes–Cooper (BCC) model

Duální (obalová) formulace – orientace na vstupy

$$\begin{aligned}
\min_{\theta,\lambda_j,s_i^-,s_r^+} \theta \quad &- \quad \varepsilon \left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ \right) \\
&\text{s.t.} \\
\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- \quad &= \quad \theta x_{i0}, \;\; i = 1, \ldots, m, \\
\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ \quad &= \quad y_{r0}, \;\; r = 1, \ldots, s, \\
\boldsymbol{\sum_{j=1}^{n} \lambda_j} \quad &= \quad 1, \\
\lambda_j \quad &\geq \quad 0, \;\; s_i^- \geq 0, \;\; s_r^+ \geq 0.
\end{aligned}$$

Příklad – 3 firmy

	1	2	3
Suroviny (vstupy)	3	6	3
Výrobky (výstupy)	7	11	5

Položme $\varepsilon = 0$

- **CRS eficientní**: firma 1

  f1: $\theta^* = 1$
  f2: $\theta^* = 0.786$
  f3: $\theta^* = 0.714$

- **VRS eficientní**: firmy 1 a 2

  f1: $\theta^* = 1$
  f2: $\theta^* = 1$
  f3: $\theta^* = ?$

Příklad – 3 firmy

	1	2	3
Suroviny (vstupy)	3	6	3
Výrobky (výstupy)	7	11	5

Položme $\varepsilon = 0$

- **CRS eficientní**: firma 1

  f1: $\theta^* = 1$
  f2: $\theta^* = 0.786$
  f3: $\theta^* = 0.714$

- **VRS eficientní**: firmy 1 a 2 i 3

  f1: $\theta^* = 1$
  f2: $\theta^* = 1$
  f3: $\theta^* = 1$ (vstupy už nejdou zlepšit $\rightarrow$ model orientovaný na výstupy)

Banker–Charnes–Cooper (BCC) model

Duální (obalová) formulace – orientace na výstupy

$$\begin{aligned}
\max_{\varphi,\lambda_j,s_i^-,s_r^+} \varphi \quad & + \quad \varepsilon\left(\sum_{i=1}^m s_i^- + \sum_{r=1}^s s_r^+\right) \\
& \text{s.t.} \\
\sum_{j=1}^n \lambda_j x_{ij} + s_i^- \quad & = \quad x_{i0},\ i=1,\dots,m, \\
\sum_{j=1}^n \lambda_j y_{rj} - s_r^+ \quad & = \quad \varphi y_{r0},\ r=1,\dots,s, \\
\sum_{j=1}^n \lambda_j \quad & = \quad 1, \\
\lambda_j \quad & \geq \quad 0,\ s_i^- \geq 0,\ s_r^+ \geq 0.
\end{aligned}$$

Tone (2001) slack-based model

$$\min_{\lambda_j, s_i^-, s_r^+} \frac{1 - 1/m \sum_{i=1}^m s_i^- / x_{i0}}{1 + 1/s \sum_{r=1}^s s_r^+ / y_{r0}}$$

$$\text{s.t.}$$

$$\begin{aligned}
\sum_{j=1}^n \lambda_j x_{ij} + s_i^- &= x_{i0},\ i = 1, \dots, m, \\
\sum_{j=1}^n \lambda_j y_{rj} - s_r^+ &= y_{r0},\ r = 1, \dots, s, \\
\lambda_j &\geq 0,\ s_i^- \geq 0,\ s_r^+ \geq 0.
\end{aligned}$$

Charnesova–Cooperova transformace na LP (na cvičení).

Další DEA modely

- Rozšíření pro finanční aplikace (diversification-consistent DEA)
- Dynamické/víceperiodické modely (dynamic/network DEA)
- Náhodná data (chance-constrained DEA)
- ...

Literatura

- Banker, R.D., Charnes, A., Cooper, W. (1984). Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis. *Management Science* 30 (9), 1078–1092.
- Charnes, A., Cooper, W. (1962). Programming with linear fractional functionals. *Naval Research Logistics Quarterly* 9, 181–196.
- Charnes, A., Cooper, W., Rhodes, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operations Research* 2, 429–444.
- Cooper, W.W., Seiford, L.M., Zhu, J. (2011). *Handbook on data envelopment analysis*, Springer, New York.
- Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operations Research* 130, 498–509.