Data Envelopment Analysis (Analýza obalu dat)

Martin Branda

Univerzita Karlova
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

Vyšetřujeme eficienci **homogenních jednotek** (Decision Making Units, DMU), $j = 1, \ldots, n$, na základě

- **Vstupů**: $X = \{x_{ij}\}$, $i = 1, \ldots, m$, preferujeme nižší hodnoty,
- **Výstupů**: $Y = \{y_{rj}\}$, $r = 1, \ldots, s$, preferujeme vyšší hodnoty.

Předpokládáme, že data jsou kladná.

Příklad – bankovní pobočky

Vstupy

- počet zaměstnanců (dále děleno dle kvalifikace: junior, senior, vedoucí)
- rozloha pobočky
- mzdové/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.)

Charnes–Cooper–Rhodes (CCR) model

Lineárně frakcionální formulace

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

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

Každá jednotka dostane pro ni nejvýhodnější váhy u_r, v_i.

Jednotka 0 je **eficientní**, právě když je optimální hodnota rovna jedné, a **neeficientní**, právě když je optimální hodnota menší než jedna.

[1] Používáme tradiční značení z DEA literatury, které však není úplně přehledné.

Charnesova–Cooperova transformace

Úlohu je možné převést na LP pomocí Charnesovy–Cooperovy transformace. 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 ..

LP po Charnesově–Cooperově transformaci:

$$\max_{u_r, v_i} \sum_{r=1}^{s} u_r y_{r0}$$

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

$$\begin{aligned}
\sum_{i=1}^{m} v_i x_{i0} &= 1, \\
\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} &\leq 0, \; j = 1, \ldots, n, \\
u_r &\geq 0, \; r = 1, \ldots, s, \\
v_i &\geq 0, \; i = 1, \ldots, m.
\end{aligned}$$

Charnes–Cooper–Rhodes (CCR) model

Duální reformulace

Dualita v LP – primární úloha:

		v_1	$\dots$	v_m	u_1	$\dots$	u_s		
		≥ 0	$\dots$	≥ 0	≥ 0	$\dots$	≥ 0		
		x_{10}	$\dots$	x_{m0}	0	$\dots$	0	$=$	1
		$-x_{11}$	$\dots$	x_{m1}	y_{11}	$\dots$	y_{s1}	$\leq$	0
			$\dots$			$\dots$			
		$-x_{1n}$	$\dots$	x_{mn}	y_{1n}	$\dots$	y_{sn}	$\leq$	0
		0	$\dots$	0	y_{10}	$\dots$	y_{s0}	max	

Charnes–Cooper–Rhodes (CCR) model
Duální reformulace

Dualita v LP – primární a duální úloha:

		v_1	$\dots$	v_m	u_1	$\dots$	u_s		
		≥ 0	$\dots$	≥ 0	≥ 0	$\dots$	≥ 0		
θ	$\in \mathbb{R}$	x_{10}	$\dots$	x_{m0}	0	$\dots$	0	$=$	1
λ_1	≥ 0	$-x_{11}$	$\dots$	x_{m1}	y_{11}	$\dots$	y_{s1}	$\leq$	0
$\vdots$		$\vdots$	$\dots$	$\vdots$	$\vdots$	$\dots$	$\vdots$		$\vdots$
λ_n	≥ 0	$-x_{1n}$	$\dots$	x_{mn}	y_{1n}	$\dots$	y_{sn}	$\leq$	0
		$\geq$		$\geq$	$\geq$		$\geq$		min
		0	$\dots$	0	y_{10}	$\dots$	y_{s0}	max	

Charnes–Cooper–Rhodes (CCR) model

Duální (obalová) formulace – orientace na vstupy

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

Interpretace: každé omezení odpovídá jednomu vstupu i, resp. výstupu r. Lineární kombinace výstupů dosahuje nejméně úrovně výstupů vyšetřované jednotky 0. Lineární kombinace vstupů dosahuje nejvýše úrovně vstupů jednotky 0 a je-li možná redukce všech vstupů, tedy $\theta < 1$, je vyšetřovaná jednotka 0 neeficientní. Není-li redukce možná, tedy $\theta = 1$, je 0 eficientní.

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

$$\max_{u_r, v_i} \sum_{r=1}^{s} u_r y_{r0}$$

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

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

Charnes–Cooper–Rhodes (CCR) model

Duální reformulace

		v_1	$\dots$	v_m	u_1	$\dots$	u_s		
		≥ 0	$\dots$	≥ 0	≥ 0	$\dots$	≥ 0		
θ	$\in \mathbb{R}$	x_{10}	$\dots$	x_{m0}	0	$\dots$	0	$=$	1
λ_1	≥ 0	$-x_{11}$	$\dots$	x_{m1}	y_{11}	$\dots$	y_{s1}	$\leq$	0
$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
λ_n	≥ 0	$-x_{1n}$	$\dots$	x_{mn}	y_{1n}	$\dots$	y_{sn}	$\leq$	0
s_1^-	≥ 0	-1	$\dots$	0	0	$\dots$	0	$\leq$	$-\varepsilon$
$\vdots$		$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$		$\vdots$
s_m^-	≥ 0	0	$\dots$	-1	0	$\dots$	0	$\leq$	$-\varepsilon$
s_1^+	≥ 0	0	$\dots$	0	-1	$\dots$	0	$\leq$	$-\varepsilon$
$\vdots$		$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$		$\vdots$
s_s^+	≥ 0	0	$\dots$	0	0	$\dots$	-1	$\leq$	$-\varepsilon$
		$\geq$		$\geq$	$\geq$		$\geq$		min
		0	$\dots$	0	y_{10}	$\dots$	y_{s0}	max	

V duální formulaci dostáváme skluzové proměnné s_i^-, s_r^+ s penalizací v účelové funkci:

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

Klasifikace jednotek

Nechť θ^*, s_i^{-*}, s_r^{+*} je optimální řešení, potom jednotka 0 je

- **Neeficientní**, pokud $\theta^* < 1$.
- **Slabě eficientní**, pokud $\theta^* = 1$ a existuje kladná skluzová proměnná, tj. $s_{\tilde{i}}^{-*} > 0$ nebo $s_{\tilde{r}}^{+*} > 0$.
- **Silně eficientní**, pokud $\theta^* = 1$ a všechny skluzové proměnné jsou nulové, tj. $s_i^{-*} = 0, \forall i$ a $s_r^{+*} = 0, \forall r$.

Production possibility set

Množina možných produktů

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

Obsahuje skutečné jednotky i umělé, které jsou vytvořeny jako nezáporné kombinace (nezáporný obal) skutečných jednotek. Vůči této množině je stanovena (ne-)eficience jednotek.

Výnosy z rozsahu

Vzniká otázka: Rostou se vstupy proporcionálně i výstupy, je to technologicky možné? Tj. platí

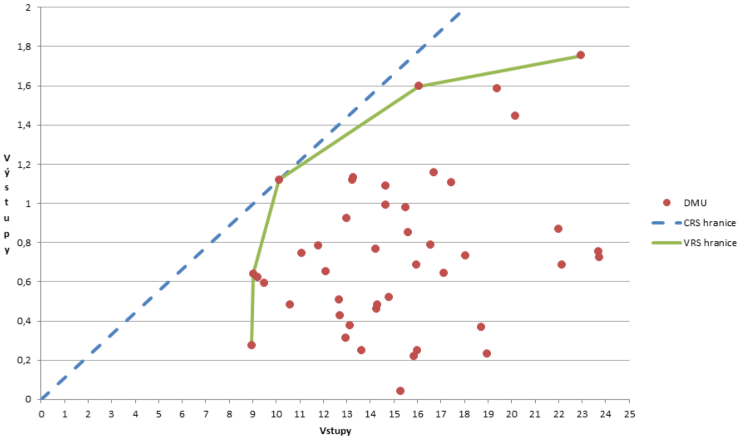
- $(x, y) \in \mathcal{PPS}$, $\alpha > 0 \Longrightarrow (\alpha x, \alpha y) \in \mathcal{PPS}$?

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

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

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

Výnosy z rozsahu



Výnosy z rozsahu

Z obrázku vidíme:

- CRS vede na nejmenší konvexní kužel obsahující data,
- VRS uvažuje horní konvexní obal dat.

Zároveň je zřejmé, že každá CRS eficientní jednotka je i VRS eficientní, naopak to ale neplatí.

Banker–Charnes–Cooper (BCC) model

Duální (obalová) formulace – orientace na vstupy

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

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

$$\max_{\varphi,\lambda_j,s_i^-,s_r^+} \varphi + \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\right)$$

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

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

Další DEA modely

Tone (2001) slack-based model = model založený (čistě) na skluzových proměnných:

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

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

Charnesova–Cooperova transformace na LP (na cvičení).

DEA ve financích

- Chceme porovnat například n investičních fondů charakterizovaných náhodnými výnosy R_j.
- Vstupy: míry rizika, transakční náklady, ...
- Výstupy: míry výnosnosti, ...

Uvažujme jednoduchý model s jedním vstupem a výstupem

$$x_{1j} = \mathrm{CVaR}_{0.95}(R_j),\ y_{1j} = \mathbb{E}[R_j].$$

Předpokládejme, že $\mathrm{CVaR}_{0.95}$ je pro naše fondy kladný[2].

[2] Tento předpoklad lze vynechat při použití obecnějších DEA modelů.

DEA ve financích

(BCC=VRS) model v duální formě

$$
\begin{aligned}
&\min_{\theta,\lambda_j} \theta \\
\text{s.t.} \quad & \sum_{j=1}^{n} \lambda_j \, \mathrm{CVaR}_{0.95}(R_j) \leq \theta \cdot \mathrm{CVaR}_{0.95}(R_0), \\
& \sum_{j=1}^{n} \lambda_j \, \mathbb{E}[R_j] \geq \mathbb{E}[R_0], \\
& \sum_{j=1}^{n} \lambda_j = 1, \\
& \lambda_j \geq 0, \; j = 1, \ldots, n.
\end{aligned}
$$

Zatímco $\sum_{j=1}^{n} \lambda_j \mathbb{E}[R_j]$ je očekávaný výnos portfolia, první omezení představuje pouze horní mez pro riziko portfolia, neboť (díky konvexitě CVaRu) dostáváme

$$\sum_{j=1}^{n} \lambda_j \, \mathrm{CVaR}_{0.95}(R_j) \geq \mathrm{CVaR}_{0.95}\left(\sum_{j=1}^{n} \lambda_j R_j\right).$$

BCC model nadhodnocuje skóre $\rightarrow$ **DEA model s diverzifikací**

$$\begin{aligned} & \min_{\theta, \lambda_j} \theta \\ \text{s.t. } \mathrm{CVaR}_{0.95}\left(\sum_{j=1}^{n} \lambda_j R_j\right) &\leq \theta \cdot \mathrm{CVaR}_{0.95}(R_0), \\ \sum_{j=1}^{n} \lambda_j \mathbb{E}[R_j] &\geq \mathbb{E}[R_0], \\ \sum_{j=1}^{n} \lambda_j &= 1, \; \lambda_j \geq 0, \; j = 1, \ldots, n. \end{aligned}$$

Další DEA modely

- Superefícience – seřazení eficientních jednotek
- Náhodná data (chance-constrained DEA)
- Záporná/reálná data (DEA based on directional-distance measure)
- Dynamické/víceperiodické modely (dynamic/network DEA)
- ...

Literatura

- Branda, M., Kopa, M.: On relations between DEA-risk models and stochastic dominance efficiency tests. *Central European Journal of Operations Research* 22 (1), 13–35.
- 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.