MODELY ANALÝZY OBALU DAT S FIXNÍM SOUČTEM VÝSTUPŮ
DATA ENVELOPMENT ANALYSIS MODELS WITH FIXED-SUM OUTPUTS

Josef Jablonský[1]

[1] prof. Ing. Josef Jablonský, CSc., Vysoká škola ekonomická v Praze, Fakulta informatiky a statistiky, Katedra ekonometrie, jablon@vse.cz, ORCID 0000-0003-0606-354X

Abstract:
Traditional data envelopment analysis (DEA) models do not consider any relations with respect to the sum of values of outputs. In many real applications, the sum of outputs is predetermined and cannot be changed. In this paper, main models considering fixed-sum outputs are formulated and discussed their properties. They usually proceed in two steps. The first step consists in deriving a new efficient frontier, usually called equilibrium efficient frontier. This frontier is computed in such a way that the values offixed-sum outputs for all units are modified to reach maximum efficiency of all units. In the second step, the efficiency score of the original units is derived with respect to the new frontier. This characteristic allows complete ranking of the units. The results of the model are illustrated on the evaluation of efficiency ofcountries attending Winter Olympic Games 2022. They are compared with traditional DEA models.

Keywords: data envelopment analysis, fixed-sum outputs, ranking, efficiency

JEL Classification: C44

t

1. ÚVOD

Tradiční modely analýzy obalu dat (DEA modely) představují nástroje pro hodnocení efektivnosti a výkonnosti souboru homogenních produkčních jednotek, které ve svých aktivitách spotřebovávají vícenásobné zdroje (vstupy) a vytvářejí vícenásobné efekty (výstupy). Uvažujme soubor n homogenních produkčních jednotek, které spotřebovávají m vstupů a produkují r výstupů. Označme x_{ij}, $i = 1, \dots, n$, $j = 1, \dots, m$, nezápornou hodnotu *j*-tého vstupu spotřebovaného i-tou jednotkou, a dále y_{ik}, $i = 1, \dots, n$, $k = 1, \dots, r$, nezápornou hodnotu *k*-tého výstupu *i*-té jednotky. V DEA modelech je míra efektivnosti i-té jednotky definována jako podíl váženého součtu výstupů a váženého součtu vstupů, tj.

$$\frac{\sum_{k=1}^{r} u_k y_{ik}}{\sum_{j=1}^{m} v_j x_{ij}} \tag{1}$$

kde u_k, $k = 1, \dots, r$ a v_j, $j = 1, \dots, m$, jsou kladné váhy jednotlivých výstupů, resp. vstupů. Základní DEA model formulovaný v (Charnes et al., 1978), často označovaný jako CCR model, maximalizuje míru efektivnosti jedné konkrétní jednotky daného souboru (označme její index jako q) za předpokladu, že míry efektivnosti všech ostatních jednotek jsou omezeny hodnotou 1. Tento optimalizační model lze tedy zapsat takto:
maximalizovat

https://doi.org/10.24132/jbt.2022.12.2.35_42

$$\theta_q = \frac{\sum_{k=1}^{r} u_k y_{qk} + \mu}{\sum_{j=1}^{m} v_j x_{qj}},$$

za podmínek (2)

$$\frac{\sum_{k=1}^{r} u_k y_{ik} + \mu}{\sum_{j=1}^{m} v_j x_{ij}} \le 1, \qquad i = 1,\ldots,n,$$

$$u_k \ge \varepsilon, v_j \ge \varepsilon, \mu = 0,$$

kde $\sum$ je infinitezimální konstanta, zajišťující pozitivitu vah, a ⌠ je proměnná, která je u CCR modelu rovna 0. Banker et al. (1984) upravili model (2) pro předpoklad variabilních výnosů z rozsahu. Pro variabilní výnosy z rozsahu je proměnná ⌠ volná (může být tedy i záporná), jinak se model (2) nemění. Často je tento model potom označovaný jako BCC model. Účelová funkce modelu (2) není lineární, ale pomocí Chanesovy-Cooperovy transformace lze snadno na lineární model převést. Podrobněji o tradičních DEA modelech např. v Dlouhý et al. (2018).

Tradiční DEA modely předpokládají, že při projekci na efektivní hranici, která je odhadnutá daným modelem, mohou být vstupy a výstupy hodnocených jednotek volně modifikovány. Uvažujeme-li ale, že všechny nebo některé výstupy modelu mají pevně daný součet v rámci celého souboru jednotek, jsou tradiční DEA modely prakticky nepoužitelné, protože projekce na efektivní hranici pevně daný součet výstupů neuvažuje. V posledních letech byla proto modelům s fixním součtem výstupů věnována značná pozornost. Yang et al. (2011) formulovali model, který minimalizuje vážený součet hodnot, o které se snižují fixní výstupy všech jednotek (kromě aktuálně hodnocené), aby se hodnocená jednotka stala efektivní. Yang et al. (2015) zavedli pojem rovnovážná efektivní hranice a formulovali model, jak této hranice dosáhnout. Tímto modelem se budeme podrobněji zabývat v následující části tohoto příspěvku. Fang (2015) a Zhu et al. (2017) koncept rovnovážné efektivní hranice upravili a rozšířili. Li et al. (2021) aplikují modely s fixním součtem výstupů pro hodnocené efektivnosti států, které se účastnily Zimních olympijských her 2018. Je evidentní, že součet výstupů (počty získaných medailí) jsou v tomto případě fixní. Oproti předchozím pracím zde autoři navíc rozšířili koncept rovnovážné efektivní hranice pro případ dvoustupňových DEA modelů.

Další část tohoto příspěvku obsahuje formulaci modelu rovnovážné efektivní hranice a její použití pro hodnocení efektivnosti pro případ výstupů s fixním součtem. Výhodou modelu je, že poskytuje kompletní uspořádání všech hodnocených jednotek. Odpadá tak nutnost aplikovat různé modely super efektivnosti, které se hojně používají při práci s tradičními DEA modely – Andersen a Petersen (1993), Tone (2002), Jablonský (2012), apod.

2. ROVNOVÁŽNÁ EFEKTIVNÍ HRANICE

Označme r_1 počet výstupů, které nejsou omezené, a r_2 počet výstupů, jejichž součet je fixní. Jsou-li hodnoty těchto výstupů z_{il}, $i = 1,\ldots, n$, $l = 1,\ldots, r_2$, potom tedy musí platit $\sum_{i=1}^{n} z_{il} = Z_l$, kde Z_l je daný součet pro l-tý výstup. Úpravou tradičního DEA modelu (2) dostáváme model pro případ, že součet některých výstupů je fixní takto:

maximalizovat

https://doi.org/10.24132/jbt.2022.12.2.35_42

$$\theta_q = \frac{\sum_{k=1}^{r_1} u_k y_{qk} + \sum_{l=1}^{r_2} w_l z_{ql} + \mu}{\sum_{j=1}^{m} v_j x_{qj}},$$

za podmínek (3)

$$\frac{\sum_{k=1}^{r_1} u_k y_{ik} + \sum_{l=1}^{r_2} w_l z_{il} + \mu}{\sum_{j=1}^{m} v_j x_{ij}} \leq 1, \qquad i = 1,\dots,n,$$

$$u_k \geq \varepsilon, v_j \geq \varepsilon, w_l \geq \varepsilon.$$

Model (Yang et al., 2015) odvozuje rovnovážnou efektivní hranici na základě minimalizace váženého součtu proměnných, které navyšují hodnoty fixních výstupů tak, aby byly všechny jednotky souboru efektivní. Model lze zapsat následovně:

minimalizovat

$$\sum_{i=1}^{n}\sum_{l=1}^{r_2} w_l \left|\delta_{il}\right|,$$

za podmínek (4)

$$\frac{\sum_{k=1}^{r_1} u_k y_{ik} + \sum_{l=1}^{r_2} w_l (z_{il} + \delta_{il}) + \mu}{\sum_{j=1}^{m} v_j x_{ij}} = 1, \qquad i = 1,\dots,n,$$

$$\sum_{i=1}^{n} \delta_{il} = 0, \qquad l = 1,\dots,r_2,$$

$$z_{il} + \delta_{il} \geq 0, \qquad i = 1,\dots,n,\ l = 1,\dots,r_2,$$

$$u_k \geq \varepsilon, v_j \geq \varepsilon, w_l \geq \varepsilon,$$

$$\delta_{il} - neomezené.$$

Proměnné ™*il* představují změnu hodnot fixních výstupů tak, aby byla *i*-tá jednotka efektivní. To zajišťuje první sada omezujících podmínek. Je zřejmé, že pro udržení fixního součtu výstupů musí být součet těchto proměnných nulový. Poslední sada podmínek zajišťuje nezápornost upravených fixních výstupů. Účelová funkce minimalizuje vážený součet absolutních hodnot ™*il* proměnných.

Model (4) je nelineární jak v účelové funkci, tak i v první sadě omezujících podmínek. Je možné jej ovšem poměrně snadno linearizovat. Stačí použít substituci $d_{il} = w_{il}\delta_{il}$. Potom účelová funkce minimalizuje součet absolutních hodnot d_{il} proměnných, což lze linearizovat použitím dvojice odchylkových proměnných, pro které už platí podmínky nezápornosti. Linearizovaný model je:

minimalizovat

$$\sum_{i=1}^{n}\sum_{l=1}^{r_2} (d_{il}^{-} + d_{il}^{+})$$

za podmínek (5)

https://doi.org/10.24132/jbt.2022.12.2.35_42

$$\sum_{k=1}^{r_1} u_k y_{ik} - \sum_{j=1}^{m} v_j x_{ij} + \sum_{l=1}^{r_2} (w_l z_{il} + d_{il}^- - d_{il}^+) + \mu = 0, \qquad i = 1,\dots,n,$$

$$\sum_{i=1}^{n} d_{il}^- - d_{il}^+ = 0, \qquad l = 1,\dots,r_2,$$

$$w_l z_{il} + d_{il}^- - d_{il}^+ \geq 0, \qquad i = 1,\dots,n,\ l = 1,\dots,r_2,$$

$$u_k \geq \varepsilon, v_j \geq \varepsilon, w_l \geq \varepsilon,$$

$$d_{il}^- \geq 0, d_{il}^+ \geq 0.$$

Po vyřešení modelu (5) lze dosazením do substitučních vztahů již snadno získat optimální hodnoty proměnných δ_{il}^*. S jejich využitím se získá míra efektivnosti q-té jednotky θ_q jako vzdálenost původních hodnot od rovnovážné efektivní hranice takto:

maximalizovat

$$\theta_q = \frac{\sum_{k=1}^{r_1} u_k y_{qk} + \sum_{l=1}^{r_2} w_l z_{ql} + \mu}{\sum_{j=1}^{m} v_j x_{qj}},$$

za podmínek (6)

$$\frac{\sum_{k=1}^{r_1} u_k y_{ik} + \sum_{l=1}^{r_2} w_l (z_{il} + \delta_{il}^*) + \mu}{\sum_{j=1}^{m} v_j x_{ij}} \leq 1, \qquad i = 1,\dots,n,$$

$$u_k \geq \varepsilon, v_j \geq \varepsilon, w_l \geq \varepsilon.$$

Při aplikaci uvedených modelů je třeba si uvědomit, že výpočet probíhá ve dvou stupních. V prvním z nich se odhaduje rovnovážná efektivní hranice řešením jedné lineární optimalizační úlohy (5). Druhá fáze řeší pro každou jednotku daného souboru jednu lineární optimalizační úlohu (6), které vede k získání míry efektivnosti jako vzdálenosti od rovnovážné efektivní hranice.

3. HODNOCENÍ STÁTŮ NA ZOH 2022

Zimní olympijské hry (ZOH) 2022 se konaly v únoru 2022 v Pekingu. Během všech soutěží bylo rozděleno 109 sad medailí (zlato, stříbro, bronz). Po skončení podobných sportovních (nejen) událostí bývají účastnické státy řazeny podle počtu získaných medailí. Pro tento účel se používá téměř výhradně lexikografické řazení, tzn. státy s vyšším počtem „lepších" medailí jsou v uspořádání výše. V celé řadě článků byly analyzovány možnosti použití DEA modelů pro uspořádání států na základě různých ukazatelů včetně aplikací pro uspořádání států podle počtu medailí na Letních či Zimních olympijských hrách. Za všechny podobné studie je možné uvést práci (Jablonský, 2018).

Tab. 1: Vstupní data

Stát	Populace	HDP	Zlato	Stříbro	Bronz
Norsko	5	366	16	8	13
Německo	82	3780	12	10	4
Čína	1370	14860	9	4	2
USA	320	20807	8	10	7
Švédsko	10	529	8	5	5
Nízozemsko	17	886	8	5	4
Rakousko	9,5	433	7	7	4
Švýcarsko	8	708	7	2	6
ROC	142	1464	6	12	14
Francie	66	2551	5	7	2
Kanada	35	1600	4	8	14
Japonsko	127	4911	3	6	9
Itálie	60	1848	2	7	8
Korea	50	1587	2	5	2
Slovinsko	2	51	2	3	2
Finsko	5,5	268	2	2	4
Nový Zéland	5	194	2	1	0
Austrálie	23	1334	1	2	1
Velká Brit.	63	2638	1	1	0
Maďarsko	10	150	1	0	2
Belgie	11	503	1	0	1
Česko	10,5	242	1	0	1
Slovensko	5,5	102	1	0	1
Bělorusko	9,5	58	0	2	0
Španělsko	47	1247	0	1	0
Ukrajina	46	142	0	1	0
Estonsko	1,5	30	0	0	1
Lotyšsko	2	33	0	0	1
Polsko	39	581	0	0	1

Zdroj: Vlastní zpracování

V naší ilustraci budeme uvažovat poslední ZOH 2022, kterých se účastnilo celkem 91 států, ale pouze 29 z nich získalo alespoň jednu medaili. Pro uspořádání podle počtu medailí budeme tedy pracovat s 29 státy (jednotkami), u kterých budou uvažovány tři výstupy – počty zlatých, stříbrných a bronzových medailí. Je zřejmé, že u všech těchto výstupů se jedná o charakteristiky s pevně daným součtem. V našem případě 109 pro každý z výstupů. Protože má každá z medailí jinou váhu, doplnili jsme modely (5) a (6) o podmínky, že váhy, odpovídající zlatým medailím jsou alespoň dvojnásobkem vah pro stříbrné medaile a ty jsou zase minimálně dvojnásobkem vah pro medaile bronzové.
Pokud bychom se chtěli ve výsledku co nejvíce přiblížit běžně používanému lexikografickému uspořádání, potom bychom zvolili DEA model bez explicitních vstupů, resp. případ, kdy všechny hodnocené jednotky mají jeden identický vstup. Pro naši ilustraci jsme ale zvolili hodnocení zemí podle jejich potenciálu, které měříme

celkovou populací daného státu (miliony obyvatel) a jeho HDP (miliardy USD). Tabulka 1 obsahuje kompletní vstupní data – státy jsou zde uspořádány lexikograficky.

Tab. 2: Výsledky

	Stát	Zlato	Stříbro	Bronz	Model (6)	Pořadí	CCR	Pořadí
1	**Norsko**	0,914	0,000	0,000	21,920	2	3,175	1
2	**Německo**	9,513	0,000	0,000	1,788	14	0,076	17
3	**Čína**	37,973	4,000	2,000	0,275	25	0,014	25
4	**USA**	18,190	67,667	7,000	0,250	26	0,011	26
5	**Švédsko**	0,000	2,655	0,000	7,918	4	0,351	4
6	**Nízozemsko**	0,000	4,449	0,000	4,725	8	0,210	10
7	**Rakousko**	0,000	2,180	0,000	9,642	3	0,410	3
8	**Švýcarsko**	0,764	2,000	0,000	5,546	6	0,273	5
9	**ROC**	3,956	0,000	0,000	3,042	11	0,135	12
10	**Francie**	5,000	2,893	0,000	1,319	16	0,054	19
11	**Kanada**	4,000	0,055	0,000	1,995	13	0,105	14
12	**Japonsko**	9,388	6,000	9,000	0,485	22	0,024	24
13	**Itálie**	1,200	7,000	0,000	1,174	17	0,054	18
14	**Korea**	2,000	4,065	0,000	1,117	18	0,044	22
15	**Slovinsko**	0,000	0,251	2,000	26,847	1	1,256	2
16	**Finsko**	0,674	0,000	0,000	4,468	9	0,221	9
17	**Nový Zéland**	0,490	0,000	0,000	5,100	7	0,236	7
18	**Austrálie**	2,340	2,000	1,000	0,599	21	0,026	23
19	**Velká Brit.**	6,153	1,000	0,000	0,225	27	0,009	27
20	**Maďarsko**	0,394	0,000	0,000	2,550	12	0,154	11
21	**Belgie**	1,000	0,527	1,000	0,792	20	0,045	21
22	**Česko**	0,622	0,000	0,000	1,612	15	0,095	16
23	**Slovensko**	0,265	0,000	0,000	3,786	10	0,224	8
24	**Bělorusko**	0,000	0,333	0,000	6,014	5	0,251	6
25	**Španělsko**	2,481	1,000	84,000	0,157	28	0,006	28
26	**Ukrajina**	0,000	0,926	0,000	1,080	19	0,051	20
27	**Estonsko**	0,075	0,000	1,000	0,433	23	0,106	13
28	**Lotyšsko**	0,084	0,000	1,000	0,394	24	0,097	15
29	**Polsko**	1,525	0,000	1,000	0,022	29	0,005	29

Zdroj: Vlastní zpracování

Tabulka 2 obsahuje výsledky získané aplikací modelů (5) a (6). Neuvažujeme-li sloupce s lexikografickým pořadím a názvem státu, jsou v dalších třech sloupcích informace o výstupech pro dosažení rovnovážné efektivní hranice – tyto souřadnice jsou získány modelem (5), doplněným o váhová omezení. V tomto případě může být oprávněná námitka, že by měly být tyto hodnoty celočíselné (počty medailí), potom by ale model (5) nemohl mít v obecném případě celočíselné přípustné řešení. Poslední dvě dvojice sloupců přinášejí míru efektivnosti vypočtenou modelem (6) resp. tradičním CCR modelem, a současně pořadí států podle této míry efektivnosti. U CCR modelu jsou u dvou států (Norsko a Slovinsko) míry super efektivnosti vypočtené Andersenovým a Petersenovým modelem. Tyto dva státy vyšly jako jediné efektivní. Výsledky nejsou překvapující. Na prvních dvou místech vyšlo s odstupem Norsko a Slovinsko. Obě země dosáhly velmi dobrých výsledků, i když v absolutních číslech Norsko mnohem lepších. Slovinsko má však nižší počet obyvatel a výrazně nižší HDP. Podle modelu s fixním součtem výstupů vychází tak Slovinsko jako první. Tradiční CCR model, resp. jeho super efektivní modifikace, upřednostňuje Norsko. Překvapivé není ani skutečnost, že státy s velkým počtem obyvatel (především Čína a USA) si oproti lexikografickému uspořádání výrazně pohoršily.

ZÁVĚR

Modely s fixním součtem výstupů představují poměrně novou kategorii modelů pro hodnocení efektivnosti a výkonnosti souboru produkčních jednotek. V článku je představený model, jehož výhoda spočívá v tom, že pro dosažení rovnovážné efektivní hranice stačí řešit jedinou lineární optimalizační úlohu. Aplikace, která je použita pro představení modelu, je spíše ilustrativní. Je zřejmé, že modely s fixním součtem výstupů mají ekonomicky zajímavější aplikace, například při rozdělování pevně daného objemu prostředků mezi podřízené jednotky.

V dalším výzkumu v této oblasti by bylo možné se zaměřit na zajímavé aplikace, směřující do ekonomické oblasti. Modely a koncept rovnovážné efektivní hranice by mohl být rozšířený o pojem α-efektivnosti, kde α je minimálně požadovaná míra efektivnosti všech jednotek souboru. V takovém případě by bylo možné uvažovat rovnovážnou efektivní hranici i pro případ celočíselných proměnných.

Poděkování

Článek je zpracovaný v rámci institucionální podpory IP400040 Fakulty Informatiky a statistiky Vysoké školy ekonomické v Praze.

ZDROJE

Andersen, P. & Petersen, N.C. (1993). A procedure for ranking efficient units in data envelopment analysis. *Management Science*, 39(10), 1261–1264.

Banker, R.D., Charnes, A. & Cooper, W.W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9), 1078–1092.

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

Dlouhý, M., Jablonský, J. & Zýková, P. (2018). *Data Envelopment Analysis* (in Czech). Praha, Professional Publishing.

Fang L (2015). A new approach for achievement of the equilibrium efficient frontier with fixed-sum outputs. *Journal of the Operational Research Society*, 67(3):412–420.

Jablonský, J. (2012). Multicriteria approaches for ranking of efficient units in DEA models. *Central European Journal of Operations Research*, 20(3), 435–449.

Jablonský, J. (2018). Ranking of countries in sporting events using two-stage data envelopment analysis models: a case of Summer Olympic Games 2016. *Central European Journal of Operations Research*. 26(4), 951–966.

Li, Y., Liu, J., Ang, S.& Yang, F. (2021). Performance evaluation of two-stage network structures with fixed-sum outputs: An application to the 2018 Winter Olympic Games. *Omega*, 102, 102342..

Tone, K. (2002). A slacks-based measure of super-efficiency in data envelopment analysis. *European Journal of Operational Research*, 143(1), 32–41.

Yang, F., Wu, D.D., Liang, L. & O'Neill, L. (2011) Competition strategy and efficiency evaluation for decision making units with fixed-sum outputs. *European Journal of Operational Research*, 212(2), 560–569.

Yang, M., Li, Y.J. & Liang, L. (2015) A generalized equilibrium efficient frontier data envelopment analysis approach for evaluation DMUs with fixed/sum outputs. *European Journal of Operational Research*, 246(1), 209–217.

Zhu, Q., Wu, J., Song, M. & Liang, L. (2017) A unique equilibrium efficient frontier with fixed sum outputs in data envelopment analysis. *Journal of the Operational Research Society*, 68(12):1483–1490.