

Analýza obalu údajov

František SUDZINA*

Úvod

Hodnotenie efektívnosti podnikov či iných druhov organizácií a procesov, v ktorých sa dajú všetky náklady a výnosy ohodnotiť trhovými cenami, je možné uskutočniť pomocou metód ekonomickej analýzy. Lenže v praxi je vhodné hodnotiť aj efektívnosť fungovania nemocníc [5], miestnych úradov či neziskových organizácií, t. j. subjektov, ktorých vstupy a výstupy sa neoceniajú v trhových cenách. V článku budeme z dôvodu zovšeobecnenia takéto ziskové či neziskové organizácie označovať ako jednotky.

Cieľom tohto článku nie je prezentácia *Data Envelopment Analysis (DEA analýzy)*, ale rozšírenie teórie o opačný obal a najhoršie koeficienty. V poslednej časti algoritimizujeme postup na identifikáciu *outliers* pomocou opačného obalu.

1. Hodnotenie efektívnosti

Existuje niekoľko metód merania efektívnosti. Zvyčajne sa však efektívnosť hodnotí mierou hospodárnosti:

$$\text{účinnosť} = \frac{\text{výstup}}{\text{vstup}} \qquad \text{náročnosť} = \frac{\text{výstup}}{\text{vstup}}$$

Takýto prístup je často neadekvátny kvôli existencii viacerých rôznorodých vstupov a výstupov. Riešenie je triviálne v prípade existencie cien. Pri ich neexistencii by sa dalo uvažovať o stanovení fiktívnych cien. Lenže na rovnakých cenách sa ťažko dohaduje, každý by si zrejme chcel stanoviť pre neho čo najvýhodnejšiu cenu na jednotlivé vstupy a výstupy.

Problematikou merania relatívnej efektívnosti s viacerými vstupmi a výstupmi sa už pred niekoľkými desaťročiami zaoberal M. J. Farrell. V príspevku uvádzané meradlo vypracoval Farrell spolu s M. Fieldhousom [6] so zameraním na konštrukciu hypotetickej efektívnej jednotky ako váženého priemeru efektívnych jednotiek, ku ktorej sa porovnáva neefektívna jednotka. Efektívna jednotka je z hľadiska Paretoho optima taká, ktorej vstupy ani výstupy sa nedajú zlepšiť bez zhoršenia iných vstupov alebo výstupov. Či je to možné, alebo nie, to sa zisťuje len

* Ing. František SUDZINA, Ekonomická univerzita v Bratislave, Podnikovohospodárska fakulta, Katedra manažmentu, Tajovského 13, 041 30 Košice

veľmi ťažko. Pri meraní relatívnej efektívnosti sa jednotka považuje za efektívnu vtedy, ak a iba ak výkon ostatných jednotiek nenasvedčuje tomu, že by sa dali zlepšiť vstupy alebo výstupy bez zhoršenia nejakého iného vstupu alebo výstupu [3].

Efektívnosť sa môže hodnotiť z dvoch hľadísk, a to ako náročnosť a účinnosť. Pri hodnotení efektívnosti z hľadiska účinnosti ide o vstupne orientovaný model. Ak je kritériom náročnosť, ide o model orientovaný na výstupy. V tejto práci bude posudzovaná relatívna efektívnosť z hľadiska účinnosti.

Bežným meradlom relatívnej efektívnosti je:

$$\text{efektívnosť} = \frac{\text{vážený súčet výstupov}}{\text{vážený súčet vstupov}}$$

čo možno zapísať ako

$$\text{efektívnosť jednotky } j = \frac{u_1 y_{1j} + u_2 y_{2j} + \dots}{v_1 x_{1j} + v_2 x_{2j} + \dots}$$

kde

u_r – váha priradená výstupu r ,

y_{rj} – množstvo výstupu r z jednotky j ,

v_i – váha priradená vstupu i ,

x_{ij} – množstvo vstupu i do jednotky j .

Relatívna efektívnosť je v tomto prístupe ohraničená intervalom $\langle 0,1 \rangle$. Ak by sa váhy nedopatrením predsa len stanovili tak, že efektívnosť niektorej jednotky by presiahla 1, alebo ak by mali všetky jednotky efektívnosť nižšiu ako 1, tak efektívnosť všetkých jednotiek sa predelí najvyššou dosiahnutou efektívnosťou. Takto sa hodnoty dostanú do intervalu $\langle 0,1 \rangle$. Počiatočným predpokladom je, že toto meradlo efektívnosti vyžaduje, aby sa aplikoval spoločný súbor váh vo všetkých jednotkách. To však nastoľuje problém, ako sa dá získať takýto dohodnutý súbor váh. V článku sa budú predpokladať konštantné výnosy z rozsahu.

2. Data Envelopment Analysis¹

Autori A. Charnes, W. Cooper a E. Rhodes [2] poznali obťažnosť hľadania spoločného súboru váh na určenie relatívnej efektívnosti. Uvedomili si oprávnenosť návrhu, že jednotky môžu mať objektívne či subjektívne dôvody na rozdielne

¹ *Data Envelopment Analysis* ešte nemá v slovenčine, podobne ako aj v češtine či v nemčine, ustálený predklad. Preto sa bude v článku používať názov tejto metódy v angličtine. Autor si je vedomý istého prekrývania pojmov, ale považuje analýzu obalu údajov za širší pojem ako *Data Envelopment Analysis*, pretože tá bola vytvorená iba na analýzu jedného z dvoch možných obalov dát.

hodnotenie vstupov a výstupov, a preto si osvoja rozličné váhy. Navrhli, aby si každá jednotka mohla zvoliť súbor váh, pri ktorých vychádza v najlepšom svetle v porovnaní s ostatnými jednotkami.

Zároveň platí, že pri uplatnení zvolených váh musí byť efektívnosť všetkých jednotiek v súbore menšia alebo rovná jednej.

Za týchto podmienok efektívnosť skúmanej jednotky j_0 sa dá získať ako riešenie nasledujúceho problému:

Maximalizuj efektívnosť jednotky j_0 , s ohľadom na efektívnosť všetkých jednotiek ≤ 1 .

Prvým krokom pri *Data Envelopment Analysis* je modelovanie skúmaného systému, to znamená určenie súvisiacich vstupov a výstupov. Je dôležité, aby sa do analýzy zahrnuli všetky vstupy, ktoré sú nápomocné pri produkcii výstupov, a zároveň všetky výstupy, ktoré sa vyrábajú zo zahrnutých vstupov. Existujú prípady, keď nie je na prvý pohľad úplne jasné, či daná premenná je vstupom, alebo je výstupom.

Niekedy je problém, z ktorého obdobia je daný údaj. Inokedy si treba presne uvedomiť, čo sa vlastne skúma.

Najlepšie je pridržať sa princípu, že vstup je taká premenná, pri ktorej platí, že čím menej sa spotrebuje, tým lepší je postup (snaha minimalizovať vstupy) a výstup je premenná, pri ktorej platí, že čím viac sa vyprodukuje, tým je lepšie (snaha maximalizovať výstupy). Na tomto mieste je nevyhnutné uviesť, že to nie je vždy tak.

Existujú aj neželané výstupy (ako napr. znečistenie prostredia), čiže výstupy, ktoré treba minimalizovať.

Ďalším rozhodnutím, ktoré sa týka výberu vstupov a výstupov, je zvolenie správnej premennej. Najčastejšie ide o rozhodovanie, ktoré prebieha medzi kvantitatívnymi a kvalitatívnymi údajmi.

Nech je n jednotiek s m vstupmi a t výstupmi, nech u je váha výstupu a v je váha vstupu, nech y je množstvo výstupu a x je množstvo vstupu. Potom algebraický model na výpočet efektívnosti jednotky j_0 vyzerá takto:

$$\text{Max } h_0 = \frac{\sum_{r=1}^t u_r y_{rj_0}}{\sum_{i=1}^m v_i x_{ij_0}}$$

$$\frac{\sum_{r=1}^t u_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}} \leq 1 \quad j = 1, 2, \dots, n$$

$$\frac{u_r}{\sum_{i=1}^m v_i x_{ij}} \geq \varepsilon \quad r = 1, 2, \dots, t$$

$$\frac{v_i}{\sum_{i=1}^m v_i x_{ij}} \geq \varepsilon \quad i = 1, 2, \dots, m$$

Premenné u a v sú ohraničené zdola. Majú byť väčšie alebo rovné ako nejaká malá kladná (nearchimedovská) veličina ε , aby sa predišlo tomu, že by sa pri určovaní efektívnosti úplne ignoroval ktorýkoľvek vstup alebo výstup. Riešenie predchádzajúceho modelu dáva hodnotu h_0 , efektívnosť jednotky j_0 a váhy vedúce k tejto efektívnosti.

Ak $h_0 = 1$, tak jednotka je efektívna relatívne k ostatným, ale ak h_0 je menšia ako 1, tak nejaká iná jednotka je efektívnejšia než skúmaná jednotka, aj keď sú váhy vybrané na maximalizovanie efektívnosti skúmanej jednotky.

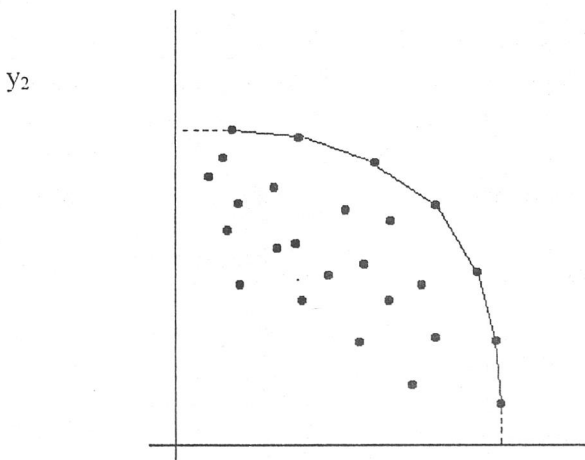
Graf 1 predstavuje príklad DEA modelu s jedným vstupom a dvoma výstupmi. Hodnoty výstupov upravené (pre jednoduchosť) na jednotkový vstup sú uvedené v tabuľke 1.

Jednotky (body) na krivke produkčných možností majú efektívnosť rovnú jednej.

Neefektívne jednotky majú efektívnosť rovnú pomeru ich vzdialenosti od začiatku k vzdialenosti ich obrazu na krivke produkčných možností od začiatku.

V tabuľke 2 sú uvedené relatívne efektívnosti skúmaných jednotiek.

Graf 1
Obal údajov



Tabuľka 1

Údaje o jednotkách

Jednotka	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
Vstup	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Výstup 1	99	72	56	80	98	39	65	19	47	75	92	58	15	30	37	66	51	80	10	30	19	14	60	17	38	
Výstup 2	11	18	30	32	30	43	43	48	51	48	51	54	64	59	60	67	70	71	80	77	72	86	84	95	92	

Tabuľka 2

Relatívna efektívnosť jednotiek

Jednotka	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Efektívnosť	1.00	0.73	0.60	0.84	1.00	0.56	0.74	0.50	0.66	0.85	1.00	0.75	0.68	0.66	0.69	0.89	0.84	1.00	0.84	0.83	0.77	0.90	1.00	1.00	1.00

Naznačený model DEA je úlohou zlomkového lineárneho programovania. Tá je ale výpočtovo náročná. Na riešenie modelu je vhodné najprv model konvertovať do lineárnej formy tak, aby sa dali použiť metódy lineárneho programovania. Lineárna verzia ohraničení je zobrazená v nasledujúcom modeli. Pre účelovú funkciu si treba všimnúť, že pri maximalizácii zlomku alebo podielu je to relatívna veľkosť čitateľa a menovateľa, o ktorú sa zaujímate, a nie ich hodnoty samotné. Možno teda dosiahnuť rovnaký efekt stanovením menovateľa rovného konštante a maximalizovaním čitateľa. Výslednou linearizáciou vznikne dvojica lineárnych modelov (model multiplikátorov):

$$\begin{aligned} \text{Min } h_0 &= \theta_0 - \left(\sum_{i=1}^m s_i^- + \sum_{r=1}^t s_r^+ \right) \\ \theta_0 x_{ij_0} - \sum_{j=1}^n x_{ij} \lambda_j - s_i^- &= 0 \\ y_{rj_0} &= \sum_{j=1}^n y_{rj} \lambda_j - s_r^+ \\ \lambda_j, s_r^+, s_i^- &\geq 0 \end{aligned}$$

a

$$\begin{aligned} \text{Max } y_0 &= \sum_{r=1}^t \mu_r y_{rj_0} \\ \sum_{i=1}^m v_i x_{ij_0} &= 1 \\ 0 &\leq \sum_{r=1}^t \mu_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 1 \\ \mu_r, v_i &\geq \varepsilon \end{aligned}$$

Na hodnotenie relatívnej efektívnosti pomocou *Data Envelopment Analysis* existujú komerčné programy, ako je napríklad *IDEAS*, *DEA Solver*, *Frontier Analyst*, *DEAP* či *Warwick DEA*. Vyznačujú sa značným užívateľským komfortom. Výpočet je možné uskutočniť aj v Exceli. Hlavne pri zlomkovom modeli si treba pri každej jednotke nastaviť nejakú konštantnú hodnotu do buniek, ktoré sa menia. Ináč sa nemusí vždy dospieť k správnejmu výsledku. Kompromisným riešením je použitie *add-inu* pre MS Excel, ktorého tvorcami sú D. K. Despotis a G. J. Smirlis [4, s. 142].

Flexibilita vo výbere váh je zároveň silnou a slabou stránkou tohto prístupu. Slabou stránkou je preto, lebo rozvážny výber váh jednotkou (možno) nesúvisiacich s hodnotou vstupu alebo výstupu môže umožniť jednotke vyzeráť efektívne, ale môže nastať pochybnosť, že to má väčšiu súvislosť s výberom váh než s akoukoľvek vlastnou efektívnosťou. Táto flexibilita je tiež silnou stránkou. Ak sa

jednotka stane neefektívnou – hoci sa použijú najvýhodnejšie váhy pri meraní jej efektívnosti – tak je to silný prejav. Potom sa najmä argument, že váhy sú nesprávne volené, nedá obhájiť.

Váhy priradené jednotlivým vstupom a výstupom neefektívnych jednotiek ukazujú na ich silné a slabé stránky. Konkrétne je možné povedať, že nulové váhy hovoria o slabých stránkach. Čím je váha vstupu či výstupu vyššia, tým silnejšiu stránku predstavuje.

Neefektívne jednotky sú neefektívne preto, lebo iná jednotka či jednotky – použitím rovnakých váh – sú efektívne. Tieto efektívne jednotky sa nazývajú referenčné jednotky.

Analýza v konečnom dôsledku slúži ako horný odhad efektívnosti transformácie vstupov na výstupy. Ak je známe množstvo vstupov, dá sa pomocou krivky zostrojenej z efektívnych jednotiek zistiť, aké bolo najväčšie množstvo výstupov, ktoré sa zatiaľ podarilo vyprodukovať, a to v rozličnej štruktúre výstupov. To síce neznamená, že viac sa nedá, ale že nevieme o žiadnej jednotke, ktorá by toho bola schopná.

Podobne, ak je známe požadované množstvo výstupov, je možné pomocou krivky zostrojenej z efektívnych jednotiek zistiť, aký najmenší počet vstupov je nevyhnutné zabezpečiť.

Takáto analýza je síce veľmi nápomocná pri tvorbe množiny produkčných možností či pri analýze jednotiek, ktorých aspoň niektoré vstupy alebo výstupy sa nedajú ohodnotiť trhovými cenami. Ale sama osebe nedáva informáciu o akejsi „absolútnej“ efektívnosti.

3. Analýza reverzného obalu údajov

Analýza reverzného obalu údajov odpovedá na otázku, či existujú také váhy, ktoré – ak by sa použili pre všetky jednotky – by spôsobili, že daná jednotka by bola najhoršia, t. j. či je možné zvoliť spoločné váhy tak, aby už žiadna iná jednotka nemohla mať nižšiu efektívnosť ako skúmaná jednotka.

Pomocou nasledujúceho modelu je možné identifikovať všetky jednotky, ktoré by mohli byť pri najhorších váhach hodnotené ako najmenej efektívne. Čím sú hodnoty bližšie 1, tým sú jednotky bližšie reverznému obalu údajov. Nižšie hodnoty poukazujú na nemožnosť stanovenia váh, pri ktorých by jednotka bola najhoršia.

$$\text{Max } h_0 = \frac{\sum_{i=1}^m v_i x_{ij_0}}{\sum_{r=1}^t u_r y_{rj_0}}$$

$$\frac{\sum_{i=1}^m v_i x_{ij}}{\sum_{r=1}^t u_r y_{rj}} \leq 1 \quad j = 1, 2, \dots, n$$

$$\frac{u_r}{\sum_{r=1}^t u_r y_{rj}} \geq \varepsilon \quad r = 1, 2, \dots, t$$

$$\frac{v_i}{\sum_{r=1}^t u_r y_{rj}} \geq \varepsilon \quad i = 1, 2, \dots, m$$

Na výpočet reverzného obalu údajov nie je potrebný nový softvér. Úplne postačuje produkt špecializovaný na *Data Envelopment Analysis*.

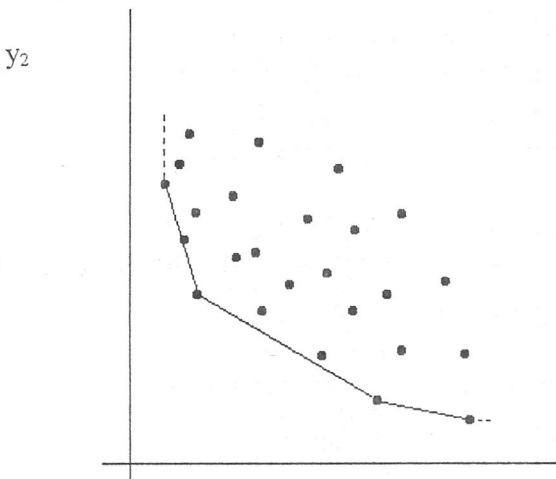
Pri výpočte reverzného obalu sa vymenia vstupy s výstupmi. Pokiaľ sú v danom programe ošetrené váhy pomocou epsilon, tak budú ošetrené aj pri výpočte reverzného obalu.

Graf 2 predstavuje príklad modelu s jedným vstupom a dvoma výstupmi. Sú v ňom použité hodnoty z tabuľky 1.

Jednotky na krivke sú najmenej efektívne a majú skóre rovné jednej. Skóre ostatných jednotiek je rovné pomeru vzdialenosti ich obrazu na krivke od začiatku k ich vzdialenosti od začiatku. Skóre všetkých jednotiek uvádzame v tabuľke 3.

Graf 2

Reverzný obal údajov



Tabuľka 3

Skóre jednotiek

Jednotka	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Skóre	1.00	1.00	0.95	0.76	0.69	0.90	0.74	1.00	0.76	0.65	0.57	0.68	0.98	0.77	0.73	0.56	0.59	0.51	1.00	0.63	0.83	0.85	0.50	0.74	0.51

Tabuľka 4

Najhoršia relatívna efektívnosť

Jednotka	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Skóre	0.11	0.18	0.30	0.32	0.30	0.39	0.43	0.19	0.47	0.48	0.52	0.55	0.15	0.30	0.37	0.67	0.52	0.72	0.10	0.30	0.19	0.14	0.61	0.17	0.38

Analýza posluží aj ako dolný odhad efektívnosti transformácie vstupov na výstupy. Ak je známe množstvo vstupov, dá sa pomocou krivky zostrojenej z najmenej efektívnych jednotiek zistiť, aké najmenšie množstvo výstupov sa zatiaľ podarilo vyprodukovať. To znamená, že síce sa dá aj viac, ale ak aj tej najhoršej jednotke postačovalo takéto množstvo, potom musí stačiť aj všetkým ostatným. Podobne, ak je známe požadované množstvo výstupov, pomocou krivky zostrojenej z najmenej efektívnych jednotiek je možné zistiť najvyšší počet vstupov, ktoré je nevyhnutné zabezpečiť.

4. Najhoršia efektívnosť

Pomocou analýzy reverzného obalu údajov sa dajú identifikovať v istom zmysle najhoršie jednotky. Ale to ešte nehovorí veľa o ich konkrétnej relatívnej efektívnosti za najhorších podmienok. Výpočet najhoršej a najlepšej možnej efektívnosti dáva lepší obraz o efektívnosti jednotiek.

Najhoršia možná relatívna efektívnosť sa dá vypočítať riešením nasledujúceho modelu:

$$\begin{aligned} \text{Min } h_0 &= \frac{\sum_{r=1}^t u_r y_{rj_0}}{\sum_{i=1}^m v_i x_{ij_0}} \\ \frac{\sum_{r=1}^t u_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}} &\leq 1 \quad j = 1, 2, \dots, n \\ \frac{u_r}{\sum_{i=1}^m v_i x_{ij}} &\geq \varepsilon \quad r = 1, 2, \dots, t \\ \frac{v_i}{\sum_{i=1}^m v_i x_{ij}} &\geq \varepsilon \quad i = 1, 2, \dots, m \end{aligned}$$

a zároveň existuje aspoň jedna jednotka, pre ktorú platí

$$\frac{\sum_{r=1}^t u_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}} = 1$$

V tomto modeli môže relatívnu efektívnosť rovnú jednej dosiahnuť najviac jedna jednotka. A to dominujúca jednotka, ak existuje. Táto jednotka je efektívna z hľadiska Paretovho optima.

Výsledky výpočtu najhoršej možnej relatívnej efektívnosti pre údaje z tabuľky 1 podľa tohto modelu sú uvedené v tabuľke 4.

Na rozdiel od analýzy reverzného obalu, tu už nie je možné použiť bežný DEA softvér, keďže to nie je úloha lineárneho programovania. Riešiteľ v Exceli nie je schopný s daným modelom pracovať, presnejšie povedané, vie nájsť iba lokálne optimum, čiže dospeje k riešeniu v závislosti od nastavenia počiatočných hodnôt. Možným riešením by bolo nainštalovanie *add-inu* na genetické algoritmy alebo linearizácia modelu.

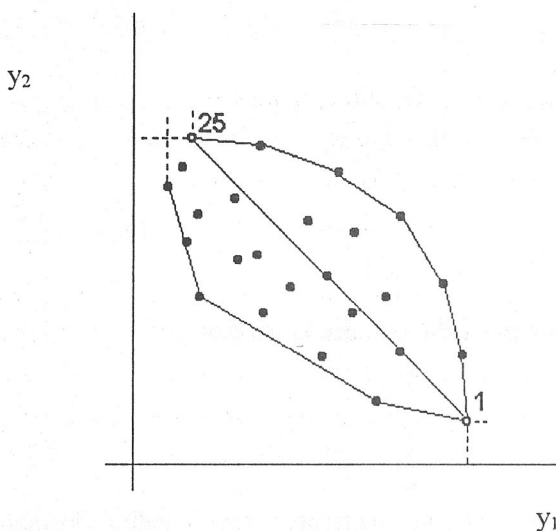
5. Outliers²

Outliers sú jednotky, ktoré ležia v extrémnych pozíciách. To znamená, že vo výstupne orientovanom modeli sú to jednotky s niektorým z najväčších výstupov, vo vstupne orientovanom modeli sú to jednotky, ktoré majú niektorý vstup značne nižší ako všetky ostatné.

Väčšinou je v nich malé zvýšenie jedného výstupu sprevádzané výrazným znížením iného výstupu, resp. malé zníženie jedného vstupu je sprevádzané výrazným zvýšením iného vstupu. Sú síce efektívne z hľadiska umiestnenia na produkčnej krivke, čiže žiadnej jednotke sa nepodarilo dosiahnuť lepšie výsledky. Ale zároveň nespĺňajú cieľ dosiahnuť efektívnosť, pretože je v nich najhoršia miera substitúcie medzi výstupmi, resp. vstupmi.

Graf 3

Outliers



² Autorovi nie je známe, či mal tento pojem v slovenčine alebo v češtine ustálený preklad. Možným vhodným výrazom by mohli byť *odľahlé jednotky*.

Všetky jednotky, ktoré ležia v oboch obaloch, sú *outliers*. Lenže nie sú to všetky *outliers*. Túto skutočnosť prezentuje graf 3. Jednotka č. 1 vpravo dole je v oboch obaloch a to je *outlier*. Jednotka č. 25 vľavo hore je *outlier*, ale nie je v oboch obaloch. Čo sa týka algoritmickej ich úplnej identifikácie, sú to jednotky, ktoré ležia na reverznom obale efektívnych jednotiek.

Outliers je možné identifikovať aj bez skúmania reverzného obalu, ale to by si vyžadovalo získať váhy vstupov a výstupov pri výpočte DEA pre každú jednotku, a následne pomocou nich upraviť vstup alebo výstup (podľa požadovaného modelu) na jednotkový a až potom by bolo možné po zložkách identifikovať, či konkrétna jednotka je, alebo nie je *outlier*.

6. Vstupy a výstupy

Vo všeobecnosti platí, že ak je t výstupov a m vstupov, dá sa očakávať $t \cdot m$ efektívnych jednotiek. Ilustrujme si to na údajoch o deviatich hypotetických jednotkách z tabuľky 1 a uvažujme len krížové pomery v tabuľke 6 (I_i znamená i -tý vstup a O_r predstavuje r -tý výstup), čiže zjednodušený prípad, keď váhy sú vyššie ako ε len pri jednom vstupe a jednom výstupe.

Tabuľka 5

Údaje o vstupoch a výstupoch hypotetických jednotiek

Jednotka	1. vstup	2. vstup	3. vstup	1. výstup	2. výstup	3. výstup
1	4	4	8	12	8	1
2	8	4	12	20	12	4
3	12	8	12	28	20	4
4	16	12	36	36	36	4
5	20	12	32	44	44	4
6	40	20	36	52	64	4
7	28	20	96	60	52	32
8	32	16	84	68	56	32
9	36	20	80	76	60	32

Tabuľka 6

Križové pomery

Jednotka	01/I1	01/I2	01/I3	02/I1	02/I2	02/I3	03/I1	03/I2	03/I3
1	3.00	3.00	1.50	2.00	2.00	1.00	0.25	0.25	0.13
2	2.50	5.00	1.67	1.50	3.00	1.00	0.50	1.00	0.33
3	2.33	3.50	2.33	1.67	2.50	1.67	0.33	0.50	0.33
4	2.25	3.00	1.00	2.25	3.00	1.00	0.25	0.33	0.11
5	2.20	3.67	1.38	2.20	3.67	1.38	0.20	0.33	0.13
6	1.30	2.60	1.44	1.60	3.20	1.78	0.10	0.20	0.11
7	2.14	3.00	0.63	1.86	2.60	0.54	1.14	1.60	0.33
8	2.13	4.25	0.81	1.75	3.50	0.67	1.00	2.00	0.38
9	2.11	3.80	0.95	1.67	3.00	0.75	0.89	1.60	0.40
Max	3.00	5.00	2.33	2.25	3.67	1.78	1.14	2.00	0.40

Maximálne hodnoty v jednotlivých stĺpcoch sú zvýraznené. Predelením každého stĺpca jeho maximálnou hodnotou dostaneme hodnotu vlastnú pre metriku DEA. Naznačuje to, že počet jednotiek v súbore má byť výrazne väčší ako $t \cdot m$, aby bol vhodný na diskrimináciu v zmysle rozlíšenia medzi efektívnosťou jednotiek.

Ďalším problémom súvisiacim s počtom vstupov a výstupov sú *outliers*. Môže ich byť $t + m$, t pre výstupne orientovaný model a m pre vstupne orientovaný model. Pravdepodobne preto sa v praxi odporúča nepoužívať viac vstupov a výstupov, ako je jedna tretina počtu jednotiek [1, s. 139].

Najvhodnejší spôsob eliminácie nejakých vstupov a výstupov je výpočet krížových korelácií osobitne medzi vstupmi a osobitne medzi výstupmi. Pokiaľ nájdeme vysokú kladnú koreláciu medzi vstupmi alebo výstupmi, ktoré navzájom súvisia, dá sa podľa uváženia vypustiť jeden z týchto ukazovateľov. Alternatívou je agregácia navzájom súvisiacich premenných. Lepšie by bolo, ak by medzi ukazovateľmi bola kladná korelácia. Ale aj bez toho je to prijateľné riešenie.

Pri stanovovaní vstupov a výstupov je nevyhnutné mať na pamäti, aby sa jednotlivé ukazovatele nachádzali v jednom vstupe alebo výstupe iba raz. Ak by vystupoval nejaký ukazovateľ oddelene, a zároveň aj ako súčasť nejakého agregovaného ukazovateľa, mal by čiastkový ukazovateľ v konečnom dôsledku minimálne takú veľkú váhu ako agregovaný ukazovateľ. Inými slovami, ostatné zložky agregovaného ukazovateľa by mohli mať najviac takú váhu ako spomínaný samostatný ukazovateľ.

V praxi môžu nastať situácie, kde spomínané tretinové pravidlo nie je možné dodržať bez toho, aby to negatívne neovplyvnilo vypovedaciu schopnosť modelu. Práve identifikácia *outliers* a ich vylúčenie zo súboru efektívnych jednotiek môže podstatne zvýšiť diskriminačnú silu analýzy obalu údajov.

V článku [3] sa uvádza, že efektívne jednotky zistené pomocou *Data Envelopment Analysis* je možné použiť ako vstup pre niektorú z klasických štatistických metód, napríklad regresiu. Elimináciou *outliers* bude mať nasledujúce štatistické spracovanie ešte vyššiu vypovedaciu schopnosť. Takto je možné urobiť akýsi odhad zhora, modelovať žiaduce správanie jednotiek. Inou možnosťou je využitie reverzného obalu údajov. Jednotky identifikované ako najmenej efektívne môžu poslúžiť pri štatistickom spracovaní na odhad zdola, ako najhoršie sledované správanie. Aj tu sa pozitívne prejaví vylúčenie *outliers*.

Záver

Cieľom predloženého príspevku bolo rozšírenie analýzy obalu údajov o opačný obal, najhoršie koeficienty a algoritmizáciu identifikácie *outliers*. V článku sme navrhli model na analýzu reverzného obalu údajov. Odpovedáme na otázku, či existujú

také váhy, ktoré by pri použití pre všetky jednotky spôsobili, že daná jednotka by bola najhoršia.

Pomocou analýzy reverzného obalu údajov sa dajú identifikovať v istom zmysle najhoršie jednotky. Ale keďže to nehovorí veľa o ich konkrétnej relatívnej efektívnosti za najhorších podmienok, bol zostavený model, ktorý vypočíta najmenšiu možnú relatívnu efektívnosť každej jednotky.

Na identifikáciu *outliers* sme navrhli použiť reverzný obal efektívnych jednotiek. Vylúčenie *outliers* zo súboru zvýši diskriminačnú silu analýzy v zmysle rozlíšenia medzi efektívnosťou jednotiek.

Došlo 4. 5. 2001

Literatúra

- [1] BANKER, R. D. – CHARNES, A. – COOPER, W. W. – SWARTS, J. – THOMAS, D. A.: An Introduction to Data Envelopment Analysis with Some of its Models and Their Uses. Research in Governmental and Nonprofit Accounting, 5, 1989.
- [2] CHARNES, A. – COOPER, W. W. – RHODES, E.: Measuring the Efficiency of Decision Making Units. European Journal of Operational Research, 2, 1978, s. 429–444.
- [3] COOPER, W. W.: Data Envelopment Analysis. In: GASS, S. I. – HARRIS, C. M.: Encyclopedia of Operations Research & Management Science. Norwell, Kluwer Academic Publishers, 1996, s. 145–150.
- [4] DESPOTIS, D. K. – SMIRLIS, G. J.: Data Envelopment Analysis with Spreadsheets. [Volume of Abstracts from 17th European Conference on Operational Research.] Budapest 2000.
- [5] DLOUHÝ, M.: An Analysis of Medical Staff Productivity. In: PLEŠINGR, J.: Mathematical Methods in Economics '99. Jindřichův Hradec, Czech Society for Operations Research, University of Economics in Prague, 1999, s. 73–77.
- [6] FARELL, M. J. – FIELDHOUSE, M.: Estimating Efficient Production Functions Under Increasing Returns to Scale. Journal of Royal Statistical Society, Series A 125, 1962, s. 252–267.

DATA ENVELOPMENT ANALYSIS

František SUDZINA

The goal of this article is to extend data envelopment analysis by reverse envelope, worst coefficients and identification of outliers.

Reverse data envelopment analysis solves the problem whether there are weights which, if applied to all units, would make the selected unit the worst efficient. In other words, whether it is possible to choose common weights, so no other units would be less efficient than the selected one.

It is possible using reverse data envelopment analysis to identify the worst units in some sense. But it does not say anything about their real relative efficiency under the worst conditions. Measurement of the worst and the best possible efficiency gives a better idea about efficiency of decision-making units.

Outliers are the units, which lie in extreme positions. These are the units that use the fewest amount in at least one input component and/or produce the highest amount in at least one output component (depending on the model orientation).

Very often is a small increase in one output compensated by a significant drop in other output or a small decrease of an input is compensated by a big increase of another one. They are efficient because they lie on a production frontier, so there is no other unit, which managed to produce better results. But in the same time, they do not fulfill the idea of being efficient because their marginal rate of substitution of inputs or outputs is the worst one.

Outliers can be identified as the ones, which lie on the reverse envelope of efficient units. So, they can be easily identified in two steps by current DEA programs.

Efficient unit identified by data envelopment analysis can be used as input for statistical (regression) analysis. By elimination of outliers my be regression analysis able to find out statistically significant information even in data where it would not be possible before. This way could the best practice be modeled. The other possibility is to use reverse data envelopment analysis and model the worst practices. Elimination of outliers may play an important role also here.