PRZEGLĄD STATYSTYCZNY
R. LVI – ZESZYT 1 – 2009

BOGUSŁAW GUZIK

PROSTA METODA DOBORU ZESTAWU NAKŁADÓW W MODELACH DEA

1. WSTĘP

W *Data Envelopment Analysis* stosowanej na przykład dla ustalania efektywności obiektów społeczno-gospodarczych, ustalania obiektów wzorcowych czy też struktury konkurencji technologicznej dokonuje się wielowymiarowego porównania wektorów nakładów poszczególnych obiektów z ich wektorami rezultatów. O ile ustalanie listy rezultatów jest stosunkowo łatwe (badacz na ogół ma wyobrażenie, co chce zbadać), o tyle ustalanie listy nakładów tak łatwe już nie jest. W literaturze dotyczącej DEA sprawę tę rozwiązuje się zazwyczaj prostym i na pewno słusznym stwierdzeniem, iż „listę nakładów należy formułować kierując się względami merytorycznymi". Naturalnie, istnieją też sugestie mające mniej dogmatyczną, a bardziej praktyczną naturę:

1. usuwanie zmiennych silnie między sobą skorelowanych: Lewin, Morry, Cook [10];
2. dołączanie zmiennych, które są silnie skorelowane z miernikiem efektywności DEA opracowanym przy węższej liście zmiennych: Norman, Stoker [11];
3. usuwanie zmiennych, które nie wpływają istotnie na informacyjność mierzoną wariancjami warunkowymi i korelacjami cząstkowymi: Jenkins, Anderson [9];
4. usuwanie zmiennych, których wykluczenie powoduje najmniej zmian współczynników efektywności, np. Wagner, Shimshak [13].

Powiedzmy zatem, że po sformułowaniu listy rezultatów, ustalono też „merytorycznie" uzasadnioną listę nakładów. Każdy, kto prowadzi badania empiryczne, nad zależnościami pewnych wielkości zawsze staje przed pytaniem, czy przyjęta lista zmiennych niezależnych, choćby „najbardziej" merytoryczna, jest dobra. Zwykle znajduje to wyraz w pytaniu, czy każda zmienna niezależna jest *istotna*, a więc czy w pierwszoplanowy (wyrazisty) sposób wpływa na kształtowanie się zmiennej zależnej (i wtedy jest istotna), czy też wpływa na nią w sposób drugo-, trzecio- a może nawet pięciorzędny (i wtedy jest nieistotna).

Problem ten ekonometria i statystyka dostrzegała od dawna i w jakiś sposób rozwiązała. Przytoczyć tu można propozycje doboru zmiennych objaśniających do modelu wykorzystujące (a) albo twierdzenia i konstrukcje statystyki matematycznej, np. badanie istotności w sensie testu t-Studenta, co jest główną ideą tzw. regresji stopniowej; (b) albo formułując pewne postulaty i wyprowadzając na tej podstawie stosowne wskaźniki jakości zestawów zmiennych niezależnych, np. metoda pojemności integralnej – Hellwig [8], metoda modeli równoważnych – Czerwiński [5] czy metoda skorygowanych współczynników determinacji – Guzik [7].

Jak można stwierdzić na podstawie literatury, w pracach empirycznych *Data Envelopment Analysis* na ogół brak tego poziomu refleksji metodologicznej, jaki jest charakterystyczny dla ekonometrii i statystyki. Dlatego problematykę doboru zmiennych objaśniających, czyli wyboru „końcowego” zestawu nakładów spośród wstępnie ustalonego spotyka się niezmiernie rzadko. Oczywiście można łatwo temu zaradzić, konstruując pewne proste metody doboru listy nakładów do modelu DEA przy ustalonej liście rezultatów. Przedkładany artykuł mieści się właśnie w tym nurcie.

Propozycja dotyczy ukierunkowanego na nakłady standardowego modelu nadefektywności CCR (*super-efficiency* CCR), kodowanego dalej przez SE-CCR[1]. Bierzemy pod uwagę ten model, a nie, na przykład, podstawowy dla DEA model CCR opracowany przez Charnesa, Coopera, Rhodesa [4], gdyż jest on ogólniejszy. W stosunku do CCR ma on ponadto wiele zalet praktycznych i analitycznych. Korzystając z niego można utworzyć ranking wszystkich obiektów, można silniej zróżnicować obiekty w pełni efektywne w sensie CCR, można przeprowadzić analizę konkurencji technologicznej. W obecnym artykule wskażemy, że na jego podstawie można także przeprowadzać analizy dotyczące wyboru „najlepszego” zestawu nakładów, czego na podstawie modelu CCR, przynajmniej w proponowany tu sposób, zrobić się nie da, z uwagi na ograniczenie wskaźnika efektywności CCR od góry do wartości 1. Dodajmy, że przytaczane na początku artykułu procedury doboru zmiennych do modelu DEA dotyczą „klasycznego” modelu CCR, a nie SE-CCR. W tym sensie sformułowana tu propozycja jest rozszerzeniem niektórych cytowanych idei, np. artykułu Jenkinsa, Andersona [9] oraz Wagner, Shimshaka [13].

Zaproponowane poniżej kryterium istotności zmiennych reprezentujących nakłady w modelach DEA opiera się na badaniu reakcji tzw. *wskaźników rankingowych* poszczególnych obiektów na zawężanie lub rozszerzanie listy nakładów. Jeśli reakcja jest duża, będziemy uważali, że odpowiednie zmienne są istotne. Jeśli natomiast zmiana jest niewielka, przyjmiemy, że odpowiednie zmienne są nieistotne. Podobnie jest też np. w ekonometrii i statystyce. Dołączenie lub usunięcie zmiennej objaśniającej istotnej w sensie testu t-Studenta, wywołuje wyraźną zmianę stopnia dopasowania i eksplancyjności modelu mierzonych współczynnikiem determinacji. Jeśli zaś zmienna jest bardzo nieistotna (np. empiryczna statystyka t nie przekracza 0,5), to usunięcie takiej zmiennej lub jej dołączenie praktycznie nie zmienia dopasowania (i eksplanacyjności).

2. MODEL SE-CCR

Analizujemy efektywność obiektu o-tego ($1 \leq o \leq J$). Pozostałe obiekty oznaczymy symbolem K_o. Można je traktować jako *konkurentów technologicznych* obiektu o–tego; $K_o = \{j = 1, \ldots, J$ z wyjątkiem $j = o\}$. Wskaźnik skuteczności (wskaźnik rankingowy) obiektu o-tego oznaczamy przez ρ_o.

Niech y_{rj} oraz x_{nj} oznaczają odpowiednio: wartość r-tego rezultatu oraz wartość n-tego nakładu w obiekcie j-tym, $j = 1, \ldots, J$; $n = 1, \ldots, N$; $r = 1, \ldots, R$ (w szczególności $j = o$). Symbolami y_{K_o} oraz x_{K_o} oznaczymy odpowiednio: wektor rezultatów oraz

[1] Za autorów modelu SE-CCR uznaje się Bankera i Gilforda [3] oraz Andersena i Petersena [1].

nakładów technologii wspólnej konkurentów obiektu o-tego. Wektory te są liniowymi kombinacjami wektorów $y_j = [y_{rj}]_{r = 1, \ldots, R}$ oraz $x_j = [x_{nj}]_{n = 1, \ldots, N}$ $(j \neq o)$, przy czym współczynniki tych kombinacji, λ_{oj}, są nieujemne:

$$y_{K_o} = \sum_{j \in K_o} \lambda_{oj} y_j \tag{1}$$

$$x_{K_o} = \sum_{j \in K_o} \lambda_{oj} x_j \tag{2}$$

Dotyczący obiektu o-tego współczynnik rankingowy, ρ_o, uzyskujemy poprzez rozwiązanie następującego zadania decyzyjnego SE-CCR:

Znaleźć takie wartości zmiennych decyzyjnych
μ_o *– mnożnik nakładów obiektu o-tego,*
$\lambda_{o1}, \ldots, \lambda_{oJ}$ *– wagi intensywności,*

że

$$\rho_o = \min \mu_o \tag{3}$$

przy warunkach:

$$y_{K_o} \geq y_o, \text{ czyli } \sum_{j \in K_o} \lambda_{oj} y_{rj} \geq y_{ro} \quad (r = 1, \ldots, R); \tag{4}$$

$$x_{K_o} \leq \mu_o x_o, \text{ czyli } \sum_{j \in K_o} \lambda_{oj} x_{nj} \leq \mu_o x_{no} \quad (n = 1, \ldots, N); \tag{5}$$

$$\mu_o, \lambda_{o1}, \ldots, \lambda_{o1} \geq 0, \lambda_{o,o} = 0 \tag{6}$$

Wskaźnik rankingowy, ρ_o, jako optymalny mnożnik nakładów, określa, jaką co najmniej krotność nakładów obiektu o-tego musiałyby ponieść pozostałe obiekty w swej optymalnej technologii wspólnej, żeby osiągnąć rezultaty uzyskane przez obiekt o-ty.

3. ZMIANY WSPÓŁCZYNNIKA RANKINGOWEGO NA SKUTEK ROZSZERZANIA (ZAWĘŻANIA) LICZBY NAKŁADÓW

Poniżej przedstawiamy najważniejsze fakty dotyczące zmian współczynników rankingowych na skutek rozszerzania lub zawężania listy nakładów lub listy rezultatów w modelu SE-CCR. Rozpatrzymy dwie sytuacje: (a) nowa lista nakładów i rezultatów powstaje przez dołączenie do wstępnej listy W pewnej innej listy B, niepustej i rozłącznej z W, (b) nowa lista powstaje przez usunięcie z listy W pewnej listy U.

Twierdzenie 1

1. Skutkiem rozszerzenia listy nakładów lub rezultatów w modelu SE-CCR jest wzrost lub stabilizacja wskaźnika rankingowego obiektu o-tego, zatem:

$$\rho_o^{W \cup B} \geq \rho_o^W, \text{ dla wszystkich } o = 1, \ldots, J \tag{7}$$

2. Skutkiem zawężenia listy nakładów lub rezultatów jest spadek lub stabilizacja wskaźnika rankingowego, czyli:

$$\rho_o^{W \setminus B} \leq \rho_o^W, \text{ dla wszystkich } o = 1, \ldots, J. \tag{8}$$

Dowód

(*Dołączenie*) Wprowadzenie dodatkowych warunków dla nakładów typu (5) lub warunków dla rezultatów typu (4) powoduje „zmniejszenie" zbioru rozwiązań dopuszczalnych zadania (3)-(6) i dlatego rozwiązanie optymalne zadania musi być *gorsze* lub co najwyżej takie samo, jak poprzednio. A ponieważ zadanie polega na minimalizacji mnożnika nakładów, dlatego w nowym zadaniu – z dodatkowymi warunkami ograniczającymi dla nakładów lub rezultatów – optymalna wartość mnożnika nakładów, czyli wskaźnik rankingowy ρ_o, musi być nie lepsza, czyli nie mniejsza od poprzedniej.

(*Usunięcie*) Usunięcie warunków dla nakładów lub rezultatów powoduje „rozluźnienie" zbioru rozwiązań dopuszczalnych, przeto nowa optymalna wartość funkcji celu musi być lepsza lub co najwyżej pozostać na poprzednim poziomie. A ponieważ zadanie polega na minimalizacji mnożnika nakładów, oznacza to spadek lub stabilizację wskaźnika rankingowego przy usunięciu nakładów lub rezultatów.

Twierdzenie 2

1. Jeśli nowa lista nakładów lub rezultatów jest połączeniem (scaleniem) dwóch rozłącznych cząstkowych list nakładów lub rezultatów, to nowy wskaźnik rankingowy jest nie mniejszy od maksimum z wskaźników rankingowych dotyczących list cząstkowych:

$$\rho_o^{W \cup B} \geq \max\left(\rho_o^W, \rho_o^B\right) \tag{9}$$

2. Jeśli natomiast nowa lista powstaje przez usunięcie niektórych nakładów lub rezultatów z poprzedniej listy, to nowy wskaźnik rankingowy jest nie mniejszy od minimum z wskaźników rankingowych dla poszczególnych składowych listy początkowej:

$$\text{jeśli } W = A \cup B, \text{ to } \rho_o^{WA}, \rho_o^{WB} \geq \min\left(\rho_o^A, \rho_o^B\right). \tag{10}$$

Dowód

(*Scalenie*) Jeśli punktem wyjścia jest lista W, to dołączenie do niej listy B skutkuje tym, że nowy wskaźnik rankingowy $\rho_o^{W \cup B} \geq \rho_o^W$. Jeśli zaś punktem wyjścia jest lista B, to nowy wskaźnik rankingowy $\rho_o^{B \cup W} \geq \rho_o^B$. A ponieważ $\rho_o^{B \cup W} = \rho_o^{W \cup B}$, więc połączenie obu tych nierówności daje nierówność (9).

(*Odłączenie*) Gdy punktem wyjścia jest lista $W = A \cup B$, to odłączenie od niej listy B skutkuje tym, że pozostaje lista A, więc nowym wskaźnikiem rankingowym będzie ρ_o^A, co na mocy (7) jest nie większe od $\rho_o^{A \cup B}$. Jeśli zaś odłączana jest lista B, to nowy wskaźnik rankingowy $\rho_o^B \leq \rho_o^{A \cup B}$. Oczywiście ρ_o^A lub ρ_o^B jest nie mniejsze od min $\left(\rho_o^A, \rho_o^B\right)$.

4. ISTOTNOŚĆ ROZSZERZENIA LUB ZAWĘŻENIA KOMBINACJI ZMIENNYCH

Nie będziemy używali standardowego dla ekonometrii i statystyki pojęcia „istotności" opartego na narzędziach statystyki matematycznej, gdyż nie czynimy tu żadnych założeń typu stochastycznego. Wydaje się zresztą, że bez „heroicznych" uproszczeń niewiele na gruncie stochastycznym da się w DEA zrobić[2]. Zastosujemy natomiast podejście opisowe. Mianowicie określimy opisowo symptomy istotności (nieistotności) nakładów, a następnie podamy procedury ustalania takiej „opisowej" istotności zmiennych charakteryzujących nakłady[3].

Zakładamy, że lista rezultatów $Y_1, \ldots, Y_R$ modelu SE-CCR oraz wyniki obserwacji wszystkich rezultatów i nakładów są ustalone.

4.1. ROZSZERZENIE KOMBINACJI

Niech dana będzie pewna „wstępna" lista nakładów W oraz zmienna D, którą chcemy dołączyć do tej listy, przy czym – oczywiście – zmienna D nie występuje w liście W. Z twierdzenia 1. wynika, iż dołączenie zmiennej D do listy W skutkować będzie wzrostem lub co najmniej stabilizacją wskaźników rankingowych $\rho_1, \ldots, \rho_J$ wszystkich badanych obiektów. Tak więc:

$$\rho_o^{W \cup D} \geq \rho_o^W, \text{ dla wszystkich } o = 1, \ldots, J. \tag{11}$$

Ponieważ dołączenie nowej zmiennej charakteryzującej nakłady nigdy, *ceteris paribus*, nie skutkuje spadkiem jakiegokolwiek wskaźnika rankingowego, zatem badanie istotności może polegać tylko na badaniu przyrostów wskaźników rankingowych (spadki, jak powiedziano, nie mogą mieć miejsca)[4].

Definicja 1

Kombinację zmiennych $W \cup D$ uznajemy za *istotną* w stosunku do kombinacji W, gdy dołączenie zmiennej D do listy W powoduje, *ceteris paribus*, *wyrazisty* wzrost wskaźników rankingowych obiektów.

---

[2] Propozycje (i to dosyć skomplikowane) stochastycznych kryteriów doboru zmiennych w modelach DEA zawierają np. prace Banker [2], Pastor, Ruiz, Sirvent [12].

[3] Dodajmy, że choć standardem współczesnej ekonometrii i statystyki są podejścia stochastyczne, podejście opisowe też jest często wykorzystywane. Np. znana metoda Z. Hellwiga doboru zmiennych objaśniających za pomocą pojemności integralnych ma taki charakter. Najpierw bowiem formułuje się symptomy „dobrych" zmiennych objaśniających (małe skorelowanie między nimi, duże ze zmienną objaśnianą), następnie formułuje się miernik odpowiadający temu postulatowi (wskaźnik integralnej pojemności informacyjnej będący sumą wskaźników informacyjności pojedynczych zmiennych objaśniających), a na koniec proponuje się wybór takich kombinacji zmiennych objaśniających, dla których wskaźnik integralny jest najlepszy.

[4] Inaczej jest w modelach regresji. Dołączenie zmiennej objaśniającej może wywołać zarówno wzrost, jak i spadek empirycznej statystyki *t*-Studenta.

Jak rozumiemy „wyrazisty” wzrost wskaźników rankingowych, powiemy za chwilę.

4.2. ZAWĘŻENIE KOMBINACJI

Z twierdzenia 2 wynika też, iż odłączenie od listy W jakiejś zmiennej U wywołuje spadek współczynnika rankingowego lub co najwyżej jego stabilizację, czyli:

$$\rho_o^{W\setminus U} \leq \rho_o^W, \text{ dla wszystkich } o = 1,\ldots, J. \tag{12}$$

Definicja 2

Kombinację zmiennych W uznajemy za istotną w stosunku do kombinacji $W\setminus U$, gdy usunięcie zmiennej U z listy W, *ceteris paribus* powoduje wyrazisty spadek wskaźników rankingowych obiektów.

4.3. SYMPTOMY ISTOTNOŚCI

Niech M oznacza pewien miernik kształtowania się wskaźników rankingowych w obiektach $o = 1, \ldots, J$. Przyjmiemy, że miernik jest tak zdefiniowany, że większa jego wartość oznacza, iż wskaźniki rankingowe są większe. W najprostszej wersji może to być suma wskaźników rankingowych wszystkich obiektów lub ich średnia.

W przypadku rozszerzania (zob. definicję 1) kombinację $W \cup D$ uznajemy za istotną w stosunku do W, gdy:

$$M_{W \cup D} \geq M_w + \Delta \tag{13}$$

lub gdy:

$$M_{W \cup D} \geq M_w (1 + \delta), \tag{14}$$

gdzie:

$\Delta > 0$ – tolerancja bezwzględna dla miernika M,

$\delta > 0$ – tolerancja względna dla miernika M (np. $\delta = 10\%$).

W przypadku zawężania (zob. definicję 2) kombinację W uznajemy za istotną w stosunku do $W\setminus U$, gdy:

$$M_{W\setminus U} \leq M_w - \Delta \tag{15}$$

lub

$$M_{W\setminus U} \leq M_w (1 - \delta). \tag{16}$$

Wartości Δ, δ w odpowiadających sobie wzorach (13), (15) oraz (14), (16) niekoniecznie muszą być takie same. Nie muszą też być równe dla różnych zmiennych.

4.4. PRZYKŁADY MIERNIKÓW

Najbardziej oczywiste jest wzięcie pod uwagę mierników *poziomu* wskaźników rankingowych, na przykład:

1. sumy (lub średniej),
2. mediany,
3. maksimum wskaźników rankingowych,
4. frakcji wskaźników większych od ustalonej liczby,
5. kwantyla (np. 90%).

Mierniki mogą też mieć inny charakter. Np. jeśli uznamy, że kombinacja zmienia się istotnie, gdy wyraźnie zmienia się *zróżnicowanie* wskaźników rankingowych, to miernikiem M mogłoby być:

6. odchylenie standardowe wskaźników rankingowych,
7. odchylenie przeciętne wskaźników rankingowych,
8. rozstęp wskaźników rankingowych.

W tym przypadku, porównując kombinację W z jej rozszerzeniem lub jej zawężeniem należy wziąć pod uwagę *moduł* różnicy między wartością M_W a wartością miernika M dla porównywanego zawężenia (rozszerzenia), bowiem miernik rozproszenia dla węższej (szerszej) listy może być zarówno większy, jak i mniejszy.

Oczywiście, każdy miernik M może być liczony na podstawie wszystkich wskaźników rankingowych lub też tylko na podstawie wybranych (= „ważnych") obiektów.

4.4. ISTOTNA KOMBINACJA ZMIENNYCH REPREZENTUJĄCYCH NAKŁADY

Powyżej podano określenia kombinacji zmiennych istotnych w stosunku do kombinacji o jeden element mniejszej. Obecnie podamy określenie kombinacji istotnej.

Definicja 3

Kombinacja W zmiennych reprezentujących nakłady jest *istotna* (przy ustalonej liście rezultatów oraz ustalonych danych empirycznych o nakładach i rezultatach), gdy jest ona istotna w stosunku do *wszystkich* swych zawężeń $W \backslash U$, gdzie $U \in X$.

Definicja 4

Za rozwiązanie zadania ustalania listy nakładów w modelu SE-CCR przyjmujemy *najliczniejszą* istotną kombinację zmiennych reprezentujących nakłady.

Podane kryterium istotności zmiennej (definicja 3) jest dość ogólne i dopuszcza, że jako kombinację istotną traktujemy taką, która zawiera podlistę nieistotną[5]. Dlatego można przyjąć, że określa ono „słaby" postulat istotności. Silniejszy postulat dotyczący istotności mógłby żądać, by istniała „ścieżka" kolejnych rozszerzeń od kombinacji jednoelementowej do kombinacji badanej taka, że wszystkie elementy na tej ścieżce są istotne.

[5] Podobnie jest zresztą w ekonometrii. Lista zmiennych objaśniających może być istotna, a jej podlista może zawierać zmienną nieistotną.

Definicja 5 („silna” istotność)

Kombinacja W jest istotna, gdy istnieje taki ciąg

$$W,\ W\backslash U_1,\ (W\backslash U_1)\backslash U_2,\ ((W\backslash U_1)\backslash U_2)\backslash U_3,\ \dots,\ U_m$$

że wszystkie wyrazy tego ciągu, są istotne względem swego zawężenia o jeden element. Kombinacja jednoelementowa z założenia jest istotna.

Dalej używamy pojęcia „słabej” istotności – zob. definicja 3.

5. PRZYKŁAD EMPIRYCZNY. PROCEDURA SEKWENCYJNEGO ZAWĘŻANIA

5.1. DANE EMPIRYCZNE

Rozpatrujemy 13 obiektów, które wstępnie scharakteryzowano dwoma rezultatami i pięcioma nakładami. Wyniki obserwacji podano w tabeli 1. Należy ustalić istotną listę nakładów dla modelu SE-CCR.

Tabela 1

Dane empiryczne

Wyszczególnienie	O_1	O_2	O_3	O_4	O_5	O_6	O_7	O_8	O_9	O_{10}	O_{11}	O_{12}	O_{13}
Y_1	9471	5859	1866	3540	17009	529	807	607	1069	680	668	6761	197
Y_2	2811	1500	1200	2000	1807	3895	535	623	118	239	16,0	1162	0,4
X_1	67,5	50,9	32,7	116,1	41,0	66,0	2,2	23,4	61,4	72,4	65,0	72,8	47,7
X_2	35,2	27,2	34,0	23,2	37,5	33,2	44,6	23,0	23,1	37,7	37,8	33,7	26,6
X_3	75,4	63,0	76,5	66,9	89,7	85,5	68,1	65,0	70,5	69,8	66,5	75,1	69,2
X_4	19836	18714	15935	19764	11077	13223	9245	3866	8172	6879	3871	1031	29,5
X_5	626,0	466,0	98,7	242,0	1367,0	57,2	5,5	4,6	13,0	12,1	7,0	549,0	6,1

Źródło: dane umowne[6].

5.2. PROCEDURA

Zamiast ogólnie o kombinacji zmiennych W, którymi mogą być nakłady lub rezultaty, od tej pory mówimy o kombinacji zmiennych. Dlatego, dla wygody, kombinację zmiennych oznaczamy przez X, a nie jak poprzednio przez W.

Przyjmiemy, że kombinacja X jest istotna względem kombinacji zawężonej o jeden element $X\backslash U$, gdy miernik M dla kombinacji zawężonej jest wyraźnie niższy od miernika M dla kombinacji X, w tym sensie, że stanowi nie więcej jak 90% miernika dla

[6] Są to dane będące kombinacją kilku przykładów, głównie danych z pracy Gospodarowicz [6].

kombinacji X. Za miernik poziomu kombinacji przyjmujemy średnią ze współczynników rankingowych:

$$M = \frac{1}{J}\sum_{o=1}^{J}\rho_o. \tag{17}$$

Istnieją trzy generalne schematy numeryczne procedury określania istotnej kombinacji zmiennych:

a) procedura sekwencyjnego zawężania,

b) procedura sekwencyjnego rozszerzania,

c) procedura kompletnego przeglądu wszystkich kombinacji zmiennych[7].

Zastosujemy schemat sekwencyjnego zawężania kombinacji wstępnej.

1. Obliczenia rozpoczynamy od najszerszego zestawu potencjalnych zmiennych reprezentujących nakłady, X, i za pomocą modelu SE-CCR wyznaczamy współczynniki rankingowe wszystkich obiektów. Liczymy też wartość miernika M_X.

2. Rozpatrujemy zawężenie kombinacji powstałe przez usunięcie zmiennej U z kombinacji X i obliczamy nowe współczynniki rankingowe wszystkich obiektów oraz wskaźnik $M_{X\backslash U}$ kombinacji $X\backslash U$. Obliczenia te wykonuje się dla wszystkich pojedynczych zmiennych U z listy X.

3. Znajdujemy tę kombinację $X\backslash U^*$, dla której wskaźnik M jest *największy*:

$$U^* = U\text{: } M_{X\backslash U^*} = \max\{M_{X\backslash U} \text{ dla wszystkich } U \in X\}. \tag{18}$$

Bierzemy maksimum, bo chodzi o to, by kombinacja X była istotna w stosunku do wszystkich swoich zawężeń o jeden element (por. definicję 3).

4. Sprawdzamy, czy kombinacja X jest istotna względem $X\backslash U^*$, czyli sprawdzamy, czy:

$$M_X \geq M_{X\backslash U^*} + \Delta \text{ lub } M_X \geq M_{X\backslash U^*}\,(1 + \delta). \tag{19}$$

5. Jeśli zajdzie (19), kombinację X uznajemy za rozwiązanie zadania doboru listy nakładów do modelu SE-CCR.

W przeciwnym przypadku przyjmujemy, że kombinacja X nie jest istotna. Wówczas listę $X\backslash U^*$, dla której wskaźnik M był minimalny, przyjmujemy za „nową" listę X, czyli algorytmicznie podstawiamy:

$$X = : X\backslash U^* \tag{20}$$

i przechodzimy do punktu 2.

[7] W terminologii doboru zmiennych objaśniających do modelu regresji odpowiada temu: procedura selekcji a posteriori, procedura selekcji a priori, procedura przeglądu kompletnego.

5.3. WYNIKI OBLICZEŃ

Etap I. $X = \{1, 2, 3, 4, 5\}$

W tabeli 2 podano współczynniki rankingowe uzyskane z ukierunkowanego na nakłady standardowego modelu SE-CCR oraz ich średnie dla kombinacji pięcioelementowej $X = \{X_1, X_2, X_3, X_4, X_5\}$ oraz jej zawężeń o jeden element.

Tabela 2

Współczynniki rankingowe (etap I)

Obiekty	X 1,2,3,4,5	Kombinacje zawężone $X\backslash U$				
		2,3,4,5	1,3,4,5	1,2,4,5	1,2,3,5	1,2,3,4
O_1	1,212	1,212	1,212	1,205	1,212	1,178
O_2	0,828	0,826	0,822	0,828	0,828	0,731
O_3	0,938	0,922	0,915	0,928	0,938	0,607
O_4	0,973	0,973	0,938	0,973	0,973	0,909
O_5	3,044	1,933	3,044	3,044	2,865	3,044
O_6	3,501	3,501	3,501	3,269	3,501	1,785
O_7	7,085	1,203	7,085	7,085	7,085	4,327
O_8	1,833	1,833	1,833	1,833	1,589	0,541
O_9	1,474	1,474	1,195	1,474	1,474	0,104
O_{10}	0,722	0,722	0,722	0,621	0,693	0,119
O_{11}	1,046	1,046	1,046	0,977	0,799	0,077
O_{12}	5,557	5,557	5,557	5,530	0,803	5,557
O_{13}	2,531	2,531	2,531	2,531	0,272	1,018
Średnia M	**2,365**	1,826	**2,339**	2,331	1,772	1,538

Źródło: obliczenia własne.

Największy wskaźnik M dla kombinacji zawężonej jest równy 2,339[8]. Jest to liczba mniejsza od 90% wartości wskaźnika dla kombinacji X, co wynosi $2{,}365 \times 0{,}9 = 2{,}128$. Dlatego kombinacji początkowej $X = \{1,2,3,4,5\}$ nie można uznać za istotną. Wzrost wskaźnika M przy przejściu od kombinacji zawężonych $X\backslash U$ do kombinacji X jest zbyt mały.

[8] Dotyczy on kombinacji {1, 3, 4, 5 }.

Etap II. $X = \{1, 3, 4, 5\}$

Wyniki obliczeń podano w tabeli 3.

Tabela 3

Współczynniki rankingowe (etap II)

Obiekty	X 1,3,4,5	Kombinacje zawężone			
		3,4,5	1,4,5	1,3,5	1,3,4
O_1	1,212	1,212	1,027	1,212	1,178
O_2	0,822	0,821	0,796	0,822	0,694
O_3	0,915	0,915	0,721	0,915	0,596
O_4	0,938	0,938	0,635	0,938	0,730
O_5	3,044	1,794	3,044	2,795	3,044
O_6	3,501	3,501	2,028	3,501	1,739
O_7	7,085	1,203	7,085	7,085	4,327
O_8	1,833	1,833	1,833	1,392	0,541
O_9	1,195	1,195	0,791	1,141	0,084
O_{10}	0,722	0,722	0,565	0,693	0,119
O_{11}	1,046	1,046	0,977	0,795	0,077
O_{12}	5,557	5,557	5,485	0,803	5,557
O_{13}	2,531	2,531	2,531	0,236	1,018
Średnia *M*	**2,339**	1,790	**2,117**	1,718	1,516

Źródło: obliczenia własne.

Ponieważ 2,117 > 2,339 × 0,90 = 2,105, dlatego kombinacji {1, 3, 4, 5} nie możemy uznać za istotną. Przechodzimy do etapu następnego, w którym $X = \{1, 4, 5\}$.

Etap III. $X = \{1, 4, 5\}$

Wyniki obliczeń podano w tabeli 4.

W tym wypadku 90% z 2,117 wynosi 1,905. Maksymalny wskaźnik M dla kombinacji zawężonych jest teraz mniejszy od tej liczby. Przejście od kombinacji {1,4} do {1,4,5} wyraźnie zwiększyło poziom współczynników rankingowych. Dlatego uznajemy, że kombinacja {1,4,5} jest istotna. Ponieważ mamy znaleźć najbardziej liczną kombinację istotną, dlatego przyjmujemy, że {1, 4, 5} jest rozwiązaniem naszego zadania.

– Badając efektywność obiektów O_1 – O_{13} ze względu na ujęte w tabeli 1 rezultaty oraz potencjalne zmienne charakteryzujące nakłady, należy więc uwzględnić trzy nakłady: X_1, X_4 oraz X_5. Nakłady X_2 oraz X_3 okazały się nieistotne.

Tabela 4

Współczynniki rankingowe (etap III)

Obiekty	X 1,4,5	Kombinacje zawężone		
		4,5	1,4	1,5
O_1	1,027	0,538	0,749	0,372
O_2	0,796	0,387	0,512	0,305
O_3	0,721	0,418	0,533	0,155
O_4	0,635	0,453	0,362	0,100
O_5	3,044	0,550	3,044	1,131
O_6	2,028	1,765	1,621	0,658
O_7	7,085	1,112	4,327	7,085
O_8	1,833	1,833	0,541	1,392
O_9	0,791	0,771	0,072	0,560
O_{10}	0,565	0,565	0,103	0,383
O_{11}	0,977	0,977	0,069	0,650
O_{12}	5,485	4,741	5,485	0,248
O_{13}	2,531	2,531	1,018	0,220
Średnia M	**2,117**	1,280	**1,418**	1,020

Źródło: obliczenia własne.

– Współczynniki rankingowe dla zadania SE-CCR z wybraną listą nakładów podaje druga kolumna tabeli 4. Wynika z niej następujące uporządkowanie obiektów:

(a) obiekty efektywne w sensie Farrella ($\rho \geq 1$):

$$(1)\ O_7,\ (2)\ O_{12},\ (3)\ O_5,\ (4)\ O_{13},\ (5)\ O_6,\ (6)\ O_8,\ (7) O_1$$

(b) obiekty nieefektywne ($\rho < 1$):

$$(8)\ O_{11},\ (9)\ O_2;\ (10)\ O_9,\ (11)\ O_3,\ (12)\ O_4,\ (13)\ O_{10}.$$

6. KRYTERIUM KOMBINOWANE

6.1. PROCEDURA

W omawianym przed chwilą przykładzie przyjmowano, że symptomem istotnej zmiany kombinacji nakładów była wyrazista zmiana jednego miernika wielkości wskaźników rankingowych. Powiedzmy jednak, że rozpatrujemy kryterium kombinowane (łączne) i badamy istotność kombinacji X względem $X\backslash U$ z uwagi na kilka mierników cząstkowych. Jeśli wartość miernika cząstkowego jest (nie jest) wystarczająca, to powiemy, że odpowiednie kryterium cząstkowe jest (nie jest) spełnione (przykłady kryteriów cząstkowych podano w 4.4.).

Definicja 6 („słabe" kryterium kombinowane)

Kombinację X uznajemy za istotną względem kombinacji $X\backslash U$, gdy spełnione jest *przynajmniej jedno* kryterium cząstkowe.

Definicja 7 („silne" kryterium kombinowane)

Kombinację X uznajemy za istotną względem kombinacji $X\backslash U$, gdy spełnione jest *każde* kryterium cząstkowe.

Definicja 8

Kombinację X uznajemy za istotną, gdy jest ona istotna względem wszystkich swoich zawężeń $X\backslash U$, $U \in X$, o jeden element.

Podobnie jak w procedurze dotyczącej pojedynczego kryterium, po ustaleniu za pomocą kryterium kombinowanego, że kombinacja X jest nieistotna, trzeba wskazać jej „najlepsze" zawężenie $X\backslash U^*$. W przypadku jednego kryterium było to proste: wybierano to zawężenie, dla którego wartość miernika M była największa. Jednakże w przypadku kryterium kombinowanego musimy liczyć się z możliwością nieidentycznych wskazań, np. jedno kryterium wskaże, iż najlepszym zawężeniem jest lista X_1, X_2, X_5 a drugie – że lista X_1, X_3, X_4.

Pojawia się wówczas problem, które wskazania zaakceptować. Naturalnie, gdyby jedno z kryteriów cząstkowych było główne, akceptowalibyśmy jego wyniki. Natomiast jeśli kryteria cząstkowe traktujemy równorzędne, trzeba przeprowadzić jakąś ich wycenę w celu ustalenia „średnio" najlepszej kombinacji $X\backslash U$.

6.2. PRZYKŁAD

Nawiązujemy do danych z tabeli 1 i chcemy ustalić „najlepszą" listę nakładów, kierując się następującym kryterium kombinowanym: Kombinację X uznajemy za istotną względem $X\backslash U$, gdy:

(a) albo średni wskaźnik rankingowy dla $X\backslash U$ stanowi mniej jak 90% średniego wskaźnika dla kombinacji X, czyli gdy $\bar{\rho}_{X\backslash U} < 0{,}9\,\bar{\rho}_X$;

(b) albo odchylenia przeciętne dla kombinacji X oraz kombinacji zawężonej różnią się przynajmniej o 10% wartości dla kombinacji X, tzn. gdy $d_{X\backslash U} < 0{,}9\ d_X$ lub $d_{X\backslash U} > 1{,}1\ d_X$,

$$d = \frac{1}{J}\sum_{o=1}^{J}\left|\rho_o - \bar{\rho}\right|; \tag{21}$$

(c) albo maksima wskaźników rankingowych w obu kombinacjach różnią się o więcej jak 10% maksimum dla kombinacji X, czyli gdy $\rho^{\max}_{X\backslash U} < 0{,}9\,\rho^{\max}_X$.

Odchylenie przeciętne po rozszerzeniu listy niekoniecznie musi wzrosnąć, dlatego kryterium (b) ma podaną postać przedziałową. Jeśli idzie o średnią oraz maksimum, to po rozszerzeniu listy ani średnia, ani maksimum nie maleje, gdyż wskaźniki rankingowe nie maleją.

Etap I. $X = \{1, 2, 3, 4, 5\}$

W tabeli 5 przedstawiono charakterystyki dla I etapu. W ostatniej kolumnie określono wynik testu: *T* oznacza, że kombinacja *X* jest istotna względem wszystkich swoich zawężeń o jeden element, *N* – że nie jest.

Tabela 5

Charakterystyki wskaźników rankingowych – etap I

Charakterystyka	*X* 1,2,3,4,5	Kombinacje zawężone *X**U*					Maksimum (minimum) miernika	90% (110%) wartości dla *X*	Wynik testu	Zawężenie *X**U**
		2,3,4,5	1,3,4,5	1,2,4,5	1,2,3,5	1,2,3,4				
Średnia $\bar{\rho}$	2,365	1,826	2,339	2,331	1,772	1,538	2,339	2,128	N (*)	1,3,4,5
Odchylenie przeciętne *d*	1,522	0,958	1,542	1,508	1,252	1,317	1,542 (0,958)	1,370 (1,674)	N (**)	1,3,4,5
Maksimum $\rho^{\max}$	7,085	5,557	7,085	7,085	7,085	5,557	7,085	6,376	N (*)	1,3,4,5

(*) – maksimum miernika większe od 90% wartości miernika dla kombinacji *X*
(**) – minimum większe od 90% miernika dla kombinacji *X* lub maksimum mniejsze od 110% miernika dla *X*

Źródło: obliczenia własne.

Etap II. $X = \{1, 3, 4, 5\}$

Tabela 6

Charakterystyki wskaźników rankingowych – etap II

Charakterystyka	*X* 1,3,4,5	Kombinacje zawężone *X**U*				Maksimum (minimum) miernika	90% (110%) wartości dla *X*	Wynik testu	Zawężenie *X**U**
		3,4,5	1,4,5	1,3,5	1,3,4				
Średnia $\bar{\rho}$	2,339	1,790	2,117	1,718	1,516	2,117	2,105	N (*)	1,4,5
Odchylenie przeciętne *d*	1,542	0,964	1,489	1,266	1,324	1,489 (0,964)	1,388 (1,697)	N (**)	1,4,5
Maksimum $\rho^{\max}$	7,085	5,557	7,085	7,085	5,557	7,085	6,376	N (*)	1,4,5

Źródło: obliczenia własne.

Etap III. $X = \{1, 4, 5\}$

Tabela 7

Charakterystyki wskaźników rankingowych – etap III

Charakterystyka	X 1,4,5	Kombinacje zawężone $X \backslash U$			Maksimum (minimum) miernika	90% (110%) wartości dla X	Wynik testu
		4,5	1,4	1,5			
Średnia $\bar{\rho}$	2,117	1,280	1,418	1,020	1,418	1,905	T
Odchylenie przeciętne d	1,489	0,885	1,355	1,007	1,355 (0,885)	1,340 (1,638)	N (**)
Maksimum $\rho^{\max}$	7,085	4,741	5,485	7,085	7,085	6,376	N (*)

(*) – maksimum miernika większe od 90% wartości miernika dla kombinacji X
(**) – minimum większe od 90% miernika dla X lub maksimum mniejsze od 110% miernika dla X

Rozwiązanie jest takie jak poprzednio: poszukiwaną kombinacją nakładów jest $\{X_1, X_4, X_5\}$, co, oczywiście, ogólnie nie musi mieć miejsca.

7. PODSUMOWANIE

1. W artykule opisano problem doboru zmiennych charakteryzujących nakłady do modelu DEA oraz zaproponowano proste kryteria badania istotności zmiennych reprezentujących nakłady w modelach Data Envelopment Analysis.

2. Kryterium ma charakter niestochastyczny i opiera się na badaniu zmian wskaźników rankingowych modelu nadefektywności DEA na zawężanie lub rozszerzanie listy nakładów. Jeśli reakcja jest duża, sądzimy, że odpowiednia zmienna (lub grupa zmiennych) jest istotna. Jeśli natomiast zmiana jest niewielka, przyjmuje się, że odpowiednie zmienne są nieistotne.

3. Charakterystyki zmian wskaźnika rankingowego mogą dotyczyć na przykład jego poziomu lub dyspersji, lub obu tych rzeczy łącznie (kryteria łączone). Za główną charakterystykę uznano maksimum różnic wskaźników rankingowych.

4. Na pewno interesujące byłoby skonstruowanie stochastycznych kryteriów doboru zmiennych do modelu DEA, np. analogicznych do powszechnie stosowanej w ekonometrii metody regresji krokowej. Wydaje się to jednak znacznie trudniejsze ze względu na bardziej zaawansowaną strukturę matematyczną rozwiązania zadania DEA (jest to rozwiązanie optymalne zadania programowania liniowego, a więc zadania optymalizacji warunkowej).

5. W artykule rozpatrzono sprawę doboru zmiennych reprezentujących nakłady, bo wydaje się, że w praktyce jest ona najczęstsza. Lista rezultatów, w sensie których badana jest efektywność obiektów gospodarczych, na ogół jest bowiem z góry zadana, a problem polega na ustaleniu listy istotnych (ważnych) nakładów, pozwalających uzyskiwać te rezultaty. Niemniej niekiedy spotyka się sytuacje odwrotne lub nawet przy-

padki jednoczesnego ustalania listy istotnych nakładów oraz rezultatów. Prezentowane w artykule idee mogą znaleźć w tych wypadkach bezpośrednie zastosowanie.

Uniwersytet Ekonomiczny w Poznaniu

LITERATURA

[1] Andersen P., Petersen N.C., [1993], *A procedure for ranking efficient units in Data Envelopment Analysis*, „Management Science", 39 (10).
[2] Banker R.D., [1993], *Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation*, „Management Science", 39 (10).
[3] Banker R.D., Gilford J.L., [1988], *A relative efficiency model for the evaluation of public health nurse productivity*, Mellon University Mimeo, Cornegie.
[4] Charnes A., Cooper W.W., Rhodes E., [1978], *Measuring the efficiency of decision making units*, „European Journal of Operational Research", 2.
[5] Czerwiński Z., [1976], *Przyczynek do dyskusji nad problemem „dobrego" modelu ekonometrycznego*, „Przegląd Statystyczny", 4.
[6] Gospodarowicz M., [2000], *Procedury analizy i oceny banków*, „Materiały i Studia", zeszyt 103, NBP, Warszawa.
[7] Guzik B., [1979], *Propozycja kryterium zmodyfikowanego współczynnika determinacji dla doboru zmiennych objaśniających do modelu ekonometrycznego*, „Przegląd Statystyczny", 1-2.
[8] Hellwig Z., [1960], *Problem optymalnego wyboru predykant*, „Przegląd Statystyczny" 3.
[9] Jenkins L., Anderson M., [2003], *A multivariate statistical approach to reducing the number of variables in data envelopment analysis*, „European Journal of Operational Research", 147 (1).
[10] Lewin A.Y., Morey R.C., Cook T.J., [1982], *Evaluating the administrative efficiency of courts*, „Omega" 10 (4).
[11] Norman M., Stoker B., [1991], *Data Envelopment Analysis: The assessment of performance*, John Wiley and Sons, Chichester.
[12] Pastor J.T., Ruiz J.L., Sirvent I., [2002], *A statistical test for nested radial DEA models*, „Operations Research", 50 (4).
[13] Wagner J.M., Shimshak D.G., [2007], *Stepwise selection of variables in data envelopment analysis: Procedures and managerial perspectives*, „European Journal of Operational Research", 180 (1).

Praca wpłynęła do redakcji we wrześniu 2008 r.

PROSTA METODA DOBORU ZESTAWU NAKŁADÓW W MODELACH DEA

Streszczenie

W artykule omówiono problem doboru zmiennych charakteryzujących nakłady do modelu DEA. Zaproponowano kryterium istotności zmiennych reprezentujących nakłady w modelach Data Envelopment Analysis opierające się na badaniu reakcji wskaźników rankingowych modelu SE-CCR na zawężanie lub rozszerzanie listy nakładów. Jeśli reakcja jest duża, uważa się, że odpowiednie zmienne są istotne. Jeśli natomiast zmiana jest niewielka, przyjmuje się, że odpowiednie zmienne są nieistotne. Podano przykłady charakterystyk omawianej reakcji. Za główną uznano maksimum różnic wskaźników rankingowych. Omówiono sprawę zmian wskaźników rankingowych przy rozszerzaniu (zawężaniu) listy obiektów.

Słowa kluczowe: DEA, dobór zmiennych, istotność nakładu/rezultatu

SIMPLE METHOD FOR INPUT SELECTION IN DEA MODELS

Summary

The article concerns the problem of selecting inputs in DEA models. The author suggests the input significance condition for Data Envelopment Analysis models which is based on the response of rank indicators in SE-CCR model to reducing or extending the list of inputs. The author assumes that a given input is significant if this response is considerable and isn't significant if the reaction is inconsiderable. Examples of the response characteristics are provided. Maximum of rank indicator differences is the main characteristic. The author discusses the problem of changes of rank indicators when extending (reducing) the list of inputs.

Key words: DEA, variable selection, input/output significance