

Artur Prędko

Uniwersytet Szczeciński

**PROPOZYCJA OPISU NIEPEWNOŚCI
W RAMACH METOD DEA I FDH***

Streszczenie: Referat jest wstępną próbą modelowania i analizy niepewności na gruncie metody DEA. Jest to jedno z możliwych rozszerzeń DEA o kierunku stochastycznym. Wspomniana niepewność dotyczy zarówno zebranych danych, jak i wartości miar efektywności technicznej obliczanych na podstawie tych właśnie danych. Zdefiniowano model statystyczny, dzięki czemu możliwe stanie się zdefiniowanie pewnych estymatorów nieparametrycznych miary efektywności technicznej. Podano własności tych estymatorów oraz ilustrujący przykład empiryczny oparty na danych rzeczywistych z polskiego sektora energetycznego.

Słowa kluczowe: metoda DEA, metoda FDH, estymacja nieparametryczna

1. Wstęp

Pierwsze prace dotyczące metody DEA (*data envelopment analysis*) pojawiły się już pod koniec lat siedemdziesiątych XX wieku (np. praca źródłowa [Charnes, Cooper, Rhodes 1978]). Od tego czasu do dzisiaj następuje na świecie ciągły rozwój zarówno zastosowań, jak i samej metodologii¹. Na gruncie polskim też pojawiło się kilka opracowań z tego zakresu (zob. np. [Gospodarowicz 2000; Kopczewski 1999; Rogowski 1996]). Najczęściej dotyczą one jednak tylko deterministycznej wersji metody i jej podstawowego zastosowania, jakim jest pomiar efektywności technicznej jednostek gospodarczych. Brakuje poszerzonej analizy zarówno różnych wersji metody, jak i związanych z tym innych zastosowań. W dotychczasowych pracach autor niniejszej pracy, chcąc wypełnić tę lukę, skoncentrował się na wykorzystaniu metody do pomiaru różnego typu efektywności obiektów (zob. np. [Prędko 2002; 2003; 2004]).

* Praca wykonana w ramach Badań Statutowych finansowanych przez Uniwersytet Ekonomiczny w Krakowie w 2008 roku.

¹ Zob. np. [Seiford 1996], gdzie znajduje się bogata bibliografia DEA do 1995 roku. Warto też zajrzeć do przeglądowej pracy [Cooper, Seiford, Tone 2006], skąd można dowiedzieć się więcej o rozwoju metody oraz znaleźć nowsze pozycje literaturowe.

Obecna praca jest pierwszą z cyklu, którego celem będzie omówienie zmian i rozszerzeń bazy teoretycznej metody DEA, jakie nastąpiły w ostatnich trzydziestu latach.

Niniejsze opracowanie stanowi wprowadzenie do jednego z możliwych rozszerzeń DEA w kierunku stochastycznym. Wspomnianemu rozszerzeniu poddamy zarówno podstawową wersję metody DEA, jak i jej modyfikację zwaną metodą FDH (*free disposal hull*), powstałą w 1984 roku (zob. [Deprins, Simar, Tulkens 1984]). Zasada działania FDH jest podobna² do DEA, jednak w jej wersji stochastycznej uzyskano dużo więcej konkretnych wyników, które warto przedstawić.

2. Oznaczenia i podstawowe definicje

Rozważamy grupę n jednostek produkcyjnych przedstawionych w postaci (x_i, y_i) ($i = 1, \dots, n$), które z p nakładów wytwarzają q produktów. Wektory $x_i = [x_i^1, \dots, x_i^p] \in \mathbb{R}_+^p$ oraz $y_i = [y_i^1, \dots, y_i^q] \in \mathbb{R}_+^q$ oznaczają odpowiednio ilości nakładów i produktów i -tego obiektu. Zakładamy, że jednostki te posługują się tą samą technologią³ reprezentowaną za pomocą zbioru możliwości produkcyjnych T . Zbiór ten definiujemy w następujący sposób:

Definicja 2.1: $T = \{(x, y) \in \mathbb{R}_+^{p+q} : z\ x\ \text{można}\ \text{wyprodukować}\ y\}$.

Za pomocą zbioru T wprowadzamy dwa, kolejne pojęcia:

Definicja 2.2: Miarą efektywności technicznej względem produktów, dla obiektu (x, y) jest wielkość:

$$\theta(x, y) = \sup \{ \theta \in \mathbb{R}_+ : (x, \theta y) \in T \}.$$

Efektywny poziom produktów dla obiektu (x, y) : $y^\theta(x) = \theta(x, y) \cdot y$.

Postać miary wskazuje, że efektywność techniczna względem produktów jest tu rozumiana jako niemożność proporcjonalnego zwiększenia ilości wszystkich produktów y i jednocześnie utrzymania ilości nakładów na dotychczasowym poziomie x . Dla jednostki efektywnej $\theta(x, y) = 1$.

Zbiór $\mathcal{X}_n = \{(x_i, y_i), i = 1, \dots, n\}$ nazwiemy **próbą** obserwowanych jednostek produkcyjnych. Posłużenie się takim określeniem oznacza konieczność wprowadzenia modelu statystycznego, który opisywałby m.in., w jaki sposób dane te zostały wygenerowane. Model ten definiują poniższe założenia.

² Zob. [Deprins, Simar, Tulkens 1984; Park, Simar, Weiner 2000]. W ramach obu metod inaczej definiuje się estymator zbioru możliwości produkcyjnych T (zob. pkt 3 pracy). Ponadto w ramach metody DEA zakłada się wypukłość zbioru T . W obu metodach analogicznie definiuje się pozostałe własności zbioru możliwości produkcyjnych oraz miarę efektywności technicznej.

³ Założenie to zapewnia możliwość porównywania ze sobą tych jednostek np. pod względem poziomu efektywności technicznej. W zastosowaniach metod FDH i DEA staramy się więc wybierać grupy obiektów z tej samej branży, sektora albo będące oddziałami większej jednostki gospodarczej (np. banku).

Założenie 1: $\forall \underline{x} \geq x, \underline{y} \leq y: [(x, y) \in T \Rightarrow (\underline{x}, \underline{y}) \in T]^4$.

Założenie to nazywane jest w literaturze założeniem o swobodnej dyslokacji nakładów i produktów (z ang. *free disposal assumption*, stąd nazwa metody FDH). W przypadku stosowania metody DEA zakłada się dodatkowo wypukłość T.

Założenie 2: Próba X_n jest realizacją *iid* zmiennych losowych na T o gęstości $f(x, y)^5$.

W celu sformułowania założenia 3 skorzystamy z reprezentacji obiektu (x, y) we współrzędnych cylindrycznych $(x, \omega(y), \eta(y))$, gdzie $\omega(y) \in \mathbb{R}_+$ jest modulem wektora produktów y , natomiast $\eta(y) \in [0, \pi/2]^{q-1}$ jest odpowiednim wektorem kątów. Przy dodatkowym założeniu istnienia gęstości $f(x)$ na \mathbb{R}_+^p , $f(\eta|x)$ na $[0, \pi/2]^{q-1}$ oraz $f(\omega|\eta, x)$ na \mathbb{R}_+ można przedstawić dekompozycję łącznej gęstości:

$$f(x, \omega, \eta) = f(\omega|\eta, x) \cdot f(\eta|x) \cdot f(x)^6.$$

Założenie 3: $\forall x \geq 0 \quad \forall \eta \in [0, \pi/2]^{q-1} \quad \exists \varepsilon_1 > 0 \quad \exists \varepsilon_2 > 0 \quad \forall \omega \in [\omega(y^o(x)) - \varepsilon_2, \omega(y^o(x))]: f(\omega|\eta, x) \geq \varepsilon_1$

Warunek ten oznacza, że w pewnym otoczeniu obiektu efektywnego $(x, y^o(x))$ istnieje realna szansa pojawienia się rzeczywistej jednostki. Jednostka $(x, y^o(x))$ jest nazywana wzorcem efektywności obiektu (x, y) i nie musi być obserwowana w rzeczywistości. Założenie to jest niezbędne dla dowiedzenia zgodności pewnych estymatorów miar efektywności (zob. pkt 3 pracy).

Założenie 4: $\forall (x, y) \in \text{int}T: \theta(x, y)$ różniczkowalna ze względu na x oraz y .

Założenie to gwarantuje odpowiednią „gładkość” miary $\theta(x, y)$ oraz odpowiedniej części brzegu zbioru T, pozwalającą uzyskać określone stopnie zbieżności wspomnianych estymatorów miar efektywności.

Za pracą [Simar, Wilson 2000] będziemy nazywać zdefiniowany model statystyczny procesem generującym dane (DGP – *data generating process*) i oznaczać przez \mathcal{P} lub równoważnie $\mathcal{P}(T, f(x, y))$ albo $\mathcal{P}(T, f(x, \omega, \eta))^7$. Estymacji będzie podlegał właśnie ów nieznan proces \mathcal{P} , a tym samym m.in. dziedzina gęstości łącznej T. Jest to jeden ze sposobów na wprowadzenie stochastyki do bazy teoretycznej wspomnianych metod⁸. Tracą one przez to swój deterministyczny charakter, ale zyskują możliwość opisu niepewności. Niepewność ta dotyczy postaci zbioru możliwości produkcyjnych, a tym samym przekłada się również na miarę efektywności technicznej $\theta(x, y)$ (zob. definicja 2.2). Te dwa elementy będą więc podlegać estymacji w punkcie trzecim pracy.

⁴ W pracy wektory porównywane są „po współrzędnych” (np. zapis $\underline{x} \geq x$).

⁵ Założenie to określa ogólnie, w jaki sposób próba jest generowana.

⁶ Wprowadzono skrócony zapis: $\omega = \omega(y)$, $\eta = \eta(y)$.

⁷ W zależności od tego, czy będziemy rozważać obiekty we współrzędnych kartezjańskich, czy biegunowych. Są to dwie równoważne charakterystyki procesu.

⁸ Inną propozycję podał np. Jati Sengupta w swoich pracach [Sengupta 1987; 1990].

Przypomnijmy, że możliwość obliczenia miary efektywności technicznej była i jest podstawowym zastosowaniem FDH oraz DEA. Kluczowe więc staje się pytanie, w jaki sposób ten nieznaną parametr jest uwzględniony we wprowadzonym modelu statystycznym. Zauważmy, że z jednej strony $y^\theta(x) = \theta(x,y) \cdot y$ (definicja 2.2), a w takim razie $\omega(y^\theta(x)) = \theta(x,y) \cdot \omega(y)$. Z drugiej strony $\omega(y^\theta(x)) = \sup\{\omega \in \mathbb{R}_+ : f(\omega | \eta, x) > 0\}$. Oznacza to, że gęstość $f(\omega | \eta, x)$ na przedziale $(0, \omega(y^\theta(x))]$ indukuje gęstość $f(\theta | \eta, x)$ na $[1, \infty)$. Tym samym istnieje również gęstość łączna:

$$f(x, \theta, \eta) = f(\theta | \eta, x) \cdot f(\eta | x) \cdot f(x).$$

O ile więc znamy T, o tyle wprowadzony model \mathcal{P} może zostać scharakteryzowany równoważnie przez T i $f(x, \theta, \eta)$, co oznacza $\mathcal{P}(T, f(x, \theta, \eta))^9$.

3. „Otoczkowe” estymatory nieparametryczne

Istotną kwestią dla estymacji nieznanego procesu \mathcal{P} jest estymacja zbioru możliwości produkcyjnych T^{10} . Jak wspomniano w punkcie drugim pracy, jest on dziedziną gęstości łącznych $f(x, y)$ i $f(x, \omega, \eta)$. Znajomość przynajmniej jego estymatora jest też niezbędna dla estymacji miary $\theta(x, y)$. Na podstawie odpowiednio metody FDH i DEA¹¹ definiuje się dwie propozycje estymatora T:

Definicja 3.1: FDH – estymatorem T jest zbiór:

$$T_{\text{FDH}}(\mathcal{X}_n) = \{(x, y) \in \mathbb{R}_+^{p+q} : \exists (x_i, y_i) \in \mathcal{X}_n \ y \leq y_i, \ x \geq x_i\}.$$

DEA – estymatorem T jest zbiór:

$$T_{\text{DEA}}(\mathcal{X}_n) = \{(x, y) \in \mathbb{R}_+^{p+q} : y \leq \sum_{i=1}^n \lambda_i y_i, \ x \geq \sum_{i=1}^n \lambda_i x_i, \ \sum_{i=1}^n \lambda_i = 1, \ \forall i=1, \dots, n; \ \lambda_i \geq 0\}$$

Przy założeniu wypukłości T oczywiste jest, iż:

$$T_{\text{FDH}}(\mathcal{X}_n) \subseteq T_{\text{DEA}}(\mathcal{X}_n) \subseteq T^{12}.$$

Znając estymator T, można zdefiniować odpowiadający mu estymator $\theta(x, y)$, korzystając z definicji tej miary (zob. definicja 2.2).

⁹ Wprowadzono oznaczenie $\theta = \theta(x, y)$. Znajomość T jest tu konieczna, aby przejść na dwie pierwsze równoważne charakterystyki. Wynika to z definicji $\theta(x, y)$. Zwróćmy też uwagę, że dziedziną $f(x, \theta, \eta)$ nie jest T w przeciwieństwie do $f(x, y)$ czy $f(x, \omega, \eta)$.

¹⁰ Z ang. *non-parametric envelopment estimators*.

¹¹ W wersji deterministycznej metod są to po prostu dwie propozycje definicji zbioru możliwości produkcyjnych T oparte na dostępnych danych (zob. prace źródłowe [Charnes, Cooper, Rhodes 1978; Deprins, Simar, Tulkens 1984]).

¹² Ze względu na kierunek inkluzji mówi się tu o estymatorach obciążonych „do wewnątrz” (z ang. *inward-biased estimators* – zob. [Simar, Wilson]).

Definicja 3.2: FDH-estymatorem miary efektywności technicznej $\theta(x,y)$ jest:

$$\theta_{\text{FDH}}(x,y) = \sup \{ \theta : (x, \theta y) \in T_{\text{FDH}}(\mathcal{X}_n) \}.$$

DEA-estymatorem miary efektywności technicznej $\theta(x,y)$ jest:

$$\theta_{\text{DEA}}(x,y) = \sup \{ \theta : (x, \theta y) \in T_{\text{DEA}}(\mathcal{X}_n) \}.$$

Z powyższych inkluzji dotyczących zbiorów uzyskujemy (o ile T jest wypukły):

$$\forall (x,y) \in T_{\text{FDH}}(\mathcal{X}_n): 1 \leq \theta_{\text{FDH}}(x,y) \leq \theta_{\text{DEA}}(x,y) \leq \theta(x,y).$$

W pracach [Korostelev, Simar, Tsybalkov 1995a; 1995b] wykazano, że estymatory $T_{\text{FDH}}(\mathcal{X}_n)$, $T_{\text{DEA}}(\mathcal{X}_n)$ są zgodne oraz ustalono szybkość ich zbieżności wg rozkładu. Ponadto dowiedziono, iż:

- estymator $T_{\text{FDH}}(\mathcal{X}_n)$ jest efektywny¹³, gdy T nie jest wypukły;
- estymator $T_{\text{DEA}}(\mathcal{X}_n)$ jest efektywny, gdy T jest wypukły.

Niestety rezultaty te uzyskano tylko dla jednego produktu ($q = 1$). Dla przypadku wielowymiarowego dowiedziono natomiast, iż estymatory $\theta_{\text{FDH}}(x,y)$, $\theta_{\text{DEA}}(x,y)$ są zgodne i ustalono szybkość ich zbieżności według rozkładu (zob. [Kneip, Park, Simar 1998; Park, Simar, Weiner 2000]). Wyprowadzono też postać asymptotycznego rozkładu próbkowego różnicy $\theta(x,y) - \theta_{\text{FDH}}(x,y)$ (zob. [Park, Simar, Weiner 2000]). Dla różnicy $\theta_{\text{DEA}}(x,y) - \theta(x,y)$ również wyprowadzono postać asymptotycznego rozkładu próbkowego, ale tylko dla przypadku jednego nakładu i produktu ($p = q = 1$) (zob. [Gijbels, Mammen, Park, Simar 1999]).

4. Ilustracja wprowadzonych pojęć

Załóżmy, iż mamy w próbie pięć obiektów, które z jednego nakładu wytwarzają jeden produkt¹⁴. Odpowiednie dane, dotyczące wartości nakładu i produktu dla poszczególnych jednostek, zebrano w tab. 1.

Tabela 1. Dane dotyczące obiektów z próby

I	1	2	3	4	5
x_i	1	3	4	5	6
y_i	2	1	3	5	4

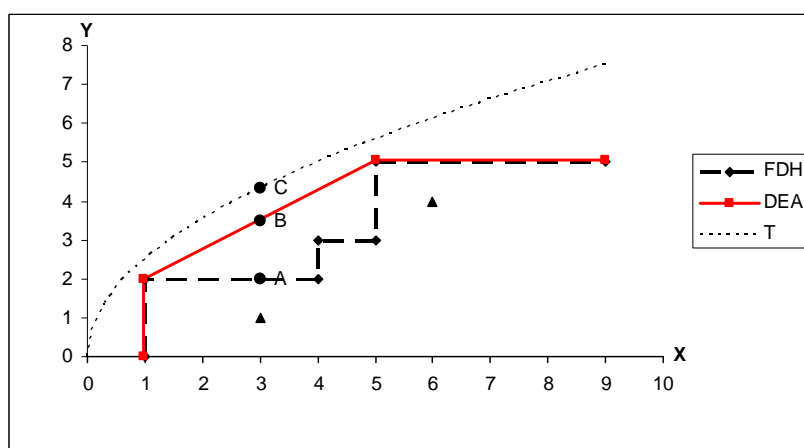
Źródło: opracowanie własne.

Zgodnie z zapisem wprowadzonym w punkcie 2 pracy, obiekt jest przedstawiany za pomocą wektora wartości jego nakładu i produktu. Na przykład w na-

¹³ Efektywny w sensie ryzyka minimaxowego (zob. [Korostelev, Simar, Tsybakov 1995a; 1995b]).

¹⁴ Umożliwi to ich ilustrację na płaszczyźnie w kartezjańskim układzie współrzędnych.

szym przypadku obiekt pierwszy $(x_1, y_1) = (1, 2)$. Tak więc naszą próbę można zapisać w następujący sposób: $\mathcal{X}_5 = \{(1,2);(3,1);(4,3);(5,5);(6,4)\}$. O jednostkach z próby zakłada się, że posługują się one tą samą technologią¹⁵. Oznacza to, że wszystkie one należą do nieznanego zbioru możliwości produkcyjnych T , który podlega estymacji (obszar ograniczony od góry drobną linią przerywaną na rys. 1). Obszary ograniczone na rysunku od góry szarą oraz czarną linią łamaną to odpowiednio zbiory $T_{DEA}(\mathcal{X}_5)$ i $T_{FDH}(\mathcal{X}_5)$. Są to postaci estymatorów zbioru T uzyskane dla naszej próby. Zauważmy, że spełnione są inkluzje wspomniane pod definicją 3.1, tzn. $T_{FDH}(\mathcal{X}_5) \subseteq T_{DEA}(\mathcal{X}_5) \subseteq T$.



Rys. 1. Ilustracja pojęć z punktu trzeciego pracy

Źródło: opracowanie własne.

Przejdźmy do ilustracji miary efektywności technicznej względem produktów oraz jej estymatorów. Weźmy dla przykładu pod uwagę obiekt $(x_2, y_2) = (3, 1)$. Jego rzutami prostopadłymi względem osi Y na granicę zbiorów $T_{FDH}(\mathcal{X}_5)$, $T_{DEA}(\mathcal{X}_5)$, T są odpowiednio obiekty $A=(a_1, a_2)=(3, 2)$, $B=(b_1, b_2)=(3, 3, 5)$ oraz $C=(c_1, c_2)=(3, c_2)$. Wartość miary efektywności i jej estymatorów dla tego obiektu wynoszą¹⁶:

$$\theta(x_2, y_2) = c_2/y_2 = c_2, \theta_{FDH}(x_2, y_2) = a_2/y_2 = 2, \theta_{DEA}(x_2, y_2) = b_2/y_2 = 3,5.$$

Jak widać, rzeczywista wartość miary dla danego obiektu (x_2, y_2) nie jest znana, ponieważ nie znamy postaci zbioru T . Możemy jedynie obliczyć wartość jej estymatora $\theta_{FDH}(x_2, y_2)$ lub $\theta_{DEA}(x_2, y_2)$ ¹⁷. Przypomnijmy jednak, że wiemy, iż:

¹⁵ Zob. początek punktu 2 pracy oraz założenie 2 modelu statystycznego.

¹⁶ Zobacz definicje 2.2 i 3.2.

¹⁷ Liczone odpowiednio na podstawie zbioru $T_{FDH}(\mathcal{X}_5)$ lub $T_{DEA}(\mathcal{X}_5)$ (zob. definicja 3.2).

$$\theta_{\text{FDH}}(x_2, y_2) \leq \theta_{\text{DEA}}(x_2, y_2) \leq \theta(x_2, y_2) \Leftrightarrow 2 \leq 3.5 \leq c_2.$$

Zwróćmy uwagę, że $\theta_{\text{FDH}}(x_1, y_1) = \theta_{\text{FDH}}(x_3, y_3) = \theta_{\text{FDH}}(x_4, y_4) = 1$, ponieważ odpowiednie obiekty leżą na brzegu zbioru $T_{\text{FDH}}(\mathcal{X}_5)$. Podobnie $\theta_{\text{DEA}}(x_1, y_1) = \theta_{\text{DEA}}(x_4, y_4) = 1$, gdyż jednostki (x_1, y_1) oraz (x_4, y_4) leżą z kolei na brzegu zbioru $T_{\text{DEA}}(\mathcal{X}_5)$. Wartości estymatorów mogą więc sugerować, iż dany obiekt jest efektywny technicznie. Nie znając jednak prawdziwej postaci zbioru T , nie możemy być tego pewni (zob. rys. 1)¹⁸. Co więcej obiekt „efektywny według estymatora θ_{FDH} ” wcale nie musi być „efektywny według estymatora θ_{DEA} ”. Przykładem jest tu obiekt (x_3, y_3) , dla którego $\theta_{\text{FDH}}(x_3, y_3) = 1 < \theta_{\text{DEA}}(x_3, y_3)$. Przy założeniu, że T jest wypukły, można w takiej sytuacji stwierdzić, iż nie jest on tak naprawdę efektywny technicznie. Są też obiekty „nieefektywne z punktu widzenia obu estymatorów” (jednostki druga i piąta). W tym przypadku możemy mieć pewność, że bez względu na postać T , nie są one rzeczywiście efektywne (o ile należą do zbioru $T_{\text{FDH}}(\mathcal{X}_5)$).

Dla zbioru T wypukłego wydaje się, że estymator $T_{\text{FDH}}(\mathcal{X}_n)$ może być wolniej zbieżny, według rozkładu, niż estymator $T_{\text{DEA}}(\mathcal{X}_n)$ (co sugeruje rys. 1). Rzeczywiście wykazano to w pracy [Kneip, Park, Simar 1998]. Jednak należy pamiętać, że dla zbioru T nie wypukłego możemy korzystać jedynie z estymatora $T_{\text{FDH}}(\mathcal{X}_n)$. Estymator $T_{\text{DEA}}(\mathcal{X}_n)$ może być w takiej sytuacji niezgodny (zob. [Simar, Wilson 2000, s. 56]).

5. Asymptotyczny rozkład próbkowy $\theta(x, y) - \theta_{\text{FDH}}(x, y)$

W pracy zajmiemy się szczegółowo jedynie asymptotycznym rozkładem próbkowym różnicy $\theta(x, y) - \theta_{\text{FDH}}(x, y)$. Inne własności, wymienione pod koniec punktu trzeciego, będą przedmiotem dalszych badań autora. Wiadomo, że same oceny punktowe miar efektywności technicznej nie wystarczą, jeśli nie powiemy nic o ich dokładności. Skoro wprowadziliśmy do metody FDH element niepewności, m.in. co do poziomu efektywności jednostki, to chcemy także mieć możliwość jego szacowania.

O postaci asymptotycznego rozkładu $\theta(x, y) - \theta_{\text{FDH}}(x, y)$ mówi twierdzenie¹⁹.

Twierdzenie 5.1: Przy założeniach 1-4 (zob. definicja modelu) dla ustalonego obiektu (x_0, y_0) :

$$\theta(x_0, y_0) - \theta_{\text{FDH}}(x_0, y_0) \sim^{\text{as.}} \text{Weibull}(n \cdot \mu^{p+q}, p+q)$$

(gdzie μ pewna nieznaną stałą zależącą od procesu \mathcal{P}).

Ważny jest również wniosek z tego twierdzenia, a mianowicie:

Wniosek 5.2: (postać asymptotycznego obciążenia i błędu standardowego $\theta_{\text{FDH}}(x_0, y_0)$)

¹⁸ Dany obiekt nie musi wcale leżeć na brzegu rzeczywistego zbioru T .

¹⁹ Dowód w pracy [Park, Simar, Weiner 2000].

$$B^{\text{as.}}(\theta_{\text{FDH}}(x_0, y_0)) = -n^{-1/(p+q)} \mu^{-1} \gamma_1$$

$$\text{ST.Dev.}^{\text{as.}}(\theta_{\text{FDH}}(x_0, y_0)) = n^{-1/(p+q)} \mu^{-1} (\gamma_2 - \gamma_1^2)^{1/2}, \text{ gdzie } \gamma_j = \Gamma\left(\frac{p+q+j}{p+q}\right), \text{ dla } j = 1, 2.$$

Zagadnienie estymacji pojawiającego się w rozkładzie asymptotycznym, nieznanego parametru μ jest dość skomplikowane, dlatego omówimy je skrótowo²⁰. Pomijamy definicję μ i przechodzimy od razu do procesu jego estymacji. Po pierwsze definiujemy zbiór:

$$\mathbf{NW}(x_0, y_0) = \{(x, y) \in \mathbb{R}_+^{p+q} : x \leq x_0, y \geq y_0\} \cap T^{21}.$$

Przypomnijmy ponadto, iż $y_0^{\hat{\circ}}(x_0) = \theta_{\text{FDH}}(x_0, y_0) \cdot y_0$. Oczywistym estymatorem tej wielkości jest więc: $\hat{y}_0^{\hat{\circ}}(x_0) = \theta_{\text{FDH}}(x_0, y_0) \cdot y_0$. Parametr μ spełnia pewien związek²², w którym kluczową rolę odgrywa prawdopodobieństwo $P[(x_i, y_i) : (x_i, y_i) \in \mathbf{NW}(x_0, y_0^{\hat{\circ}}(x_0) - zy_0)]$, gdzie z to dowolna liczba dodatnia. Estymatorem tej wielkości jest:

$$\hat{p}_{NW,0}(z) = \#\{(x_i, y_i) : (x_i, y_i) \in \mathbf{NW}(x_0, \hat{y}_0^{\hat{\circ}}(x_0) - zy_0)\} / n^{23}.$$

Na koniec niech $z = Cn^{-2/[3(p+q)]}$, gdzie C to dowolna, dodatnia stała. Zgodnym²⁴, skorygowanym o obciążenie, estymatorem μ jest wtedy:

$$\hat{\mu} = z^{-1} \hat{p}_{NW,0}(z)^{1/(p+q)} - \gamma_1 z^{-1} (n-1)^{-1/(p+q)}.$$

Na mocy wniosku 5.2 skorygowany o obciążenie estymator miary efektywności technicznej ma postać:

$$\hat{g}(x_0, y_0) = \theta_{\text{FDH}}(x_0, y_0) + \gamma_1 \hat{\mu}^{-1} n^{-1/(p+q)}.$$

Wiąże się to również z postacią asymptotycznego przedziału ufności dla parametru $\theta(x_0, y_0)$. Niech z_α oznacza α kwantyl rozkładu $W(1, p+q)$, tzn. $z_\alpha = (-\ln(1-\alpha))^{1/(p+q)}$. Wtedy otrzymujemy:

$$P[\theta_{\text{FDH}}(x_0, y_0) \leq \theta(x_0, y_0) \leq \theta_{\text{FDH}}(x_0, y_0) + \hat{\mu}^{-1} n^{-1/(p+q)} z_{1-\alpha}] \approx^{\text{as.}} 1 - \alpha.$$

Wybór jednostronnego przedziału ufności wynika stąd, iż rozkład Weibulla jest ograniczony od dołu przez zero.

²⁰ Szczegóły ponownie w pracy [Park, Simar, Weiner 2000].

²¹ Część zbioru T leżąca na pfn.-zach. od obiektu (x_0, y_0) (z ang. *NW* – *North West*).

²² $\forall z \rightarrow 0$: $P(x_i \leq x_0, y_i \geq y_0^{\hat{\circ}}(x_0) - zy_0) = (\mu z)^{p+q} + \Theta(z^{p+q+1})$. Szczegóły w pracy [Park, Simar, Weiner 2000].

²³ Na mocy zał. 2 $(x_i, y_i) \in T$, dla dowolnego $i = 1, \dots, n$, nie zachodzi więc konieczność estymacji T w definicji zbioru $\mathbf{NW}(x_0, \hat{y}_0^{\hat{\circ}}(x_0) - zy_0)$.

²⁴ Estymator $\hat{\mu}$ jest zgodny z optymalną szybkością zbieżności dla każdego ustalonego C . W praktyce, w skończonych próbach, dobiera się jednak parametr C tak, by zminimalizować błąd średniokwadratowy. Czyni się to w każdym zastosowaniu osobno na podstawie dostępnych danych.

6. Przykład empiryczny

Ilustrację empiryczną przeprowadzimy na danych rzeczywistych z 1995 r. dotyczących 32 polskich elektrowni i elektrociepłowni. Będzie to grupa jednostek produkcyjnych, których efektywność techniczną będziemy analizować. Za nakłady przyjęto:

- kapitał (wartość brutto środków trwałych liczona w zł);
- pracę (liczba pracowników);
- energię wsadu (liczoną w TJ).

Produktem działalności jednostek jest wytworzona energia (liczona w TJ²⁵). Dane, uszeregowane według malejącej wielkości produkcji, przedstawiono w tab. 2.

Tabela 2. Dane rzeczywiste dotyczące 32 obiektów z polskiego sektora energetycznego

Lp.	Kapitał (zł)	Praca (os.)	Energia wsadu (TJ)	Energia wytworzona (TJ)	Lp.	Kapitał (zł)	Praca (os.)	Energia wsadu (TJ)	Energia wytworzona (TJ)
1	4 453 039,7	5 895	274 452	100 749	17	377 291,0	1 193	37 182	14 641,1
2	769 796,6	4 364	158 639	54 677	18	225 282,5	1 254	14 444	12 752,9
3	1 408 093,0	5 463	63 939	51 499,5	19	332 627,9	1125	11 964	10 185,1
4	950 768,6	1 553	86 343	34 356,7	20	206 713,5	836	23 748	10 119,2
5	2 357 441,6	3 270	98 488	33 991,7	21	420 620,1	1 138	16 443	9 406,5
6	1 152 842,6	3 242	79 329	33 731,4	22	385 322,7	828	24 671	8 964
7	1 126 615,9	2 884	85 119	32 164,4	23	322 211,3	465	9 274	7 922,1
8	954 553,0	3 307	33 305	29 404,3	24	320 969,2	503	9 216	6 530
9	1 648 631,6	2 464	40355	28 615,7	25	138 156,0	528	12 201	4 900,7
10	1 270 206,3	2 556	78 037	28 536,7	26	191 445,3	537	7 729	4 533,9
11	547 787,3	2 273	22 711	20 160,4	27	163 589,9	567	7 208	4 472,5
12	6 666 579,9	1 640	51 163	19 731,6	28	111 331,3	526	7 257	4 424,7
13	828 724,3	1 519	48 527	17 874,2	29	94 102,4	420	4 802	3 926,6
14	589 864,9	1 249	22 219	16 620,7	30	77 688,8	320	9 316	3 320,9
15	441 469,2	790	24 115	16 428,7	31	59 697,6	637	5 313	3 153,1
16	378 370,8	1 399	32 506	15 318,7	32	65 256,9	336	2 393	1 939

Źródło: [Osiewalski, Wróbel-Rotter 2002].

Do obliczeń wartości wielkości, omówionych w poprzednich częściach pracy, użyto procedur napisanych w programie Gauss 6.0²⁶. Uzyskane wyniki przedstawiono w tab. 3.

Ze względu na estymator $\theta_{FDH}(x_i, y_i)$ większość obiektów z próby uznano za efektywne technicznie względem produktów., nieefektywne są jedynie jednostki o nr 5 i 21 ($\theta_{FDH}(x_i, y_i) > 1$, dla $i = 5, 21$, zob. czwarta kolumna tab. 3). Wynika to z za-

²⁵ 1GWh = 3,6TJ (teradzul).

²⁶ Autorką procedur jest dr Anna Pajor z Katedry Ekonometrii UEK, której składam serdeczne podziękowania za wielką pomoc okazaną przy realizacji tej pracy.

łożenia 1 modelu statystycznego o swobodnej dyslokacji nakładów i produktów. Oznacza ono bowiem, że musi zająć bardzo szczególna zależność między ilością nakładów i produktów pewnych obiektów z próby, aby jeden z nich został uznany za nieefektywny. Dotyczy to obiektów o numerach 4 i 5 oraz 21 i 19. Weźmy dla ilustracji pierwszą z tych par. Zauważmy, że ilości nakładów jednostki nr 5 są większe od ilości nakładów jednostki nr 4 (dla wszystkich trzech nakładów). Pomimo to produkcja jednostki nr 5 jest mniejsza niż produkcja jednostki nr 4 (zob. tab. 2).

Tabela 3. Wartości odpowiednich estymatorów oraz końce asymptotycznych przedziałów ufności*

I	$\hat{\mu}$	$\hat{g}(x_i, y_i)$	$L = \theta_{FDH}(x_i, y_i)$	P	L - P
1	2	3	4	5	6
1	0,301	2,264	1,000	2,835	1,835
2	0,065	6,889	1,000	9,548	8,548
3	0,206	2,846	1,000	3,679	2,679
4	0,436	1,874	1,000	2,269	1,269
5	0,680	1,571	1,011	1,824	0,813
6	0,488	1,781	1,000	2,133	1,133
7	0,534	1,714	1,000	2,036	1,036
8	0,436	1,874	1,000	2,269	1,269
9	0,534	1,714	1,000	2,036	1,036
10	0,575	1,662	1,000	1,961	0,961
11	0,375	2,016	1,000	2,475	1,475
12	0,680	1,561	1,000	1,814	0,814
13	0,680	1,561	1,000	1,814	0,814
14	0,375	2,016	1,000	2,475	1,475
15	0,206	2,846	1,000	3,679	2,679
16	0,436	1,874	1,000	2,269	1,269
17	0,436	1,874	1,000	2,269	1,269
18	0,065	6,889	1,000	9,548	8,548
19	0,436	1,874	1,000	2,269	1,269
20	0,375	2,016	1,000	2,475	1,475
21	0,375	2,099	1,083	2,558	1,475
22	0,534	1,714	1,000	2,036	1,036
23	0,206	2,846	1,000	3,679	2,679
24	0,206	2,846	1,000	3,679	2,679
25	0,375	2,016	1,000	2,475	1,475
26	0,301	2,264	1,000	2,835	1,835
27	0,206	2,846	1,000	3,679	2,679
28	0,206	2,846	1,000	3,679	2,679
29	0,206	2,846	1,000	3,679	2,679
30	0,065	6,889	1,000	9,548	8,548
31	0,065	6,889	1,000	9,548	8,548
32	0,065	6,889	1,000	9,548	8,548

* L – lewy koniec, P – prawy koniec, poziom istotności: 0,05. We wzorze na $\hat{\mu}$ przyjęto $C = 1$, za pracę [Park, Simar, Weiner 2000] (zob. przypis 24).

Źródło: opracowanie własne.

Taka sytuacja może oczywiście świadczyć o nieefektywności obiektu nr 5. Jednak w praktyce, zwłaszcza w małych próbach, trudno oczekiwać, aby taka szczególna zależność występowała często. W naszej próbie 32-elementowej są tylko dwie takie sytuacje, stąd właśnie mała liczba obiektów nieefektywnych.

Przy małych próbach, takich jak nasza, warto jednak wziąć pod uwagę obciążenie uzyskanych wyników. Spójrzmy na wartości estymatora $\theta_{\text{FDH}}(x_i, y_i)$ skorygowanego o obciążenie (zob. kolumna trzecia w tab. 3). Po pierwsze ani jeden z obiektów nie jest tu uznawany za efektywny ($\hat{\mathcal{G}}(x_i, y_i) > 1$, dla $i = 1, \dots, 32$). Po drugie wartości $\hat{\mathcal{G}}(x_i, y_i)$ są bardziej zróżnicowane i narzucają zupełnie inny ranking efektywności niż estymator $\theta_{\text{FDH}}(x_i, y_i)$. Na przykład obiekt nr 5, uprzednio nieefektywny, obecnie jest jednym z bardziej efektywnych (wartość $\hat{\mathcal{G}}(x_5, y_5)$ bliska jedności).

Wszystkie uzyskane wartości estymatorów (również te skorygowane) należy traktować z dużą ostrożnością, o czym świadczy choćby szerokość przedziałów ufności (kolumna 6 w tab. 3). Na przykład dla jednostek nr 2, 18, 30, 31, 32 mamy najmniejszą wiedzę na temat tego, ile tak naprawdę wynosi rzeczywista wartość efektywności technicznej tych obiektów; choć z drugiej strony wiemy z dużym prawdopodobieństwem, w jakim przedziale może się ona znajdować, i to dla każdej jednostki²⁷. Do ostrożności, przy interpretacji wyników w tak małej próbie jak nasza, skłaniają też jedynie asymptotyczne własności wykorzystywanych estymatorów.

Literatura

- Charnes A., Cooper W.W., Rhodes E., *Measuring the efficiency of decision making units*, "European Journal of Operations Research" 1978 vol. 2, s. 429-444.
- Cooper W.W., Seiford L.M., Tone K., *Introduction to Data Envelopment Analysis and Its Uses*, Springer, New York 2006.
- Deprins D., Simar L., Tulkens H., *Measuring labor efficiency in post offices*, [w:] *The Performance of Public Enterprises: Concepts and Measurements*, red. M. Marchand, P. Pestieau, H. Tulkens, North-Holland, Amsterdam 1984, s. 243-267.
- Gijbels I., Mammen E., Park B.U., Simar L., *On estimation of monotone and concave frontier function*, "Journal of the American Statistical Association" 1999 vol. 94, s. 220-228.
- Gospodarowicz M., *Procedury analizy i oceny banków*, Materiały i Studia NBP, Warszawa 2000.
- Kneip A., Park B.U., Simar L., *A note on the convergence of nonparametric DEA estimators for production efficiency scores*, "Econometric Theory" 1998 vol. 14, s. 783-793.
- Kopczewski T., *Racjonalność zachowań banków komercyjnych w Polsce w latach 1994-1997*, praca doktorska, Uniwersytet Warszawski, Wydział Nauk Ekonomicznych (1999).
- Korostelev A.P., Simar L., Tsybakov A.B., *Efficient estimation of monotone boundaries*, "The Annals of Statistics" 1995a vol.23, s. 476-489.
- Korostelev A.P., Simar L., Tsybakov A.B., *On estimation of monotone and convex boundaries*, Publications de l'Institut de Statistique des Universités de Paris XXXIX 1995b vol. 1, s. 3-18.

²⁷ Interpretacja nieformalna przedziału ufności.

- Osiewalski J., Wróbel-Rotter R., *Bayesowski model efektów losowych w analizie efektywności kosztowej (na przykładzie elektrowni i elektrociepłowni polskich)*, „Przegląd Statystyczny” 2002 vol. 50 nr 2, s.47-68.
- Park B.U., Simar L., Weiner Ch., *The FDH estimator for productivity efficiency scores*, “Econometric Theory” 2000 vol.16, s. 855-877.
- Prędko A., *Stale i zmienne czynniki produkcji w badaniu efektywności kosztowej za pomocą metody DEA*, „Przegląd Statystyczny” 2002 vol. 49 nr 3, s. 121-134.
- Prędko A., *Analiza efektywności za pomocą metody DEA: podstawy formalne i ilustracja ekonomiczna*, „Przegląd Statystyczny” 2003 vol. 50 nr 1, s. 87-100.
- Prędko A., *Badanie efektywności jednostek gospodarczych w sferze przychodów za pomocą metody DEA*, [w:] *Metody i zastosowania badań operacyjnych '04*, red. T. Trzaskalik, Wydawnictwo AE, Katowice 2004, s. 285-295.
- Rogowski G., *Analiza i ocena działalności banków z wykorzystaniem metody DEA*, „Bank i Kredyt” 1996 nr 9.
- Seiford L.M., *DEA: The evolution of the state of the art (1978-1995)*, “Journal of Productivity Analysis” 1996 vol.7, s. 99-137.
- Sengupta J.K., *DEA for efficiency measurement in the stochastic case*, “Computers and Operations Research” 1987 vol.14, s. 117-129.
- Sengupta J.K., *Transformations in stochastic DEA models*, “Journal of Econometrics” 1990 vol. 46, s. 109-123.
- Simar L., Wilson P.W., *Statistical inference in nonparametric frontier models: The state of the art*, “Journal of Productivity Analysis” 2000 vol. 3, s. 49-78.

THE PROPOSAL OF THE UNCERTAINTY DESCRIPTION USING THE DEA AND FDH METHODS

Summary: The article is an initial attempt to modeling and analyzing of uncertainty within the DEA and FDH methods. The uncertainty relates to both the collected data and values of the technical efficiency measure, which are calculated basing on these data. It is one of the possible extension of the DEA/FDH method in stochastic direction. The statistical model and nonparametric estimators of the technical efficiency measure are defined. Properties of these estimators and the empirical example (based on the real data) are presented.