PRZEGLĄD STATYSTYCZNY
R. LXI – ZESZYT 2 – 2014

ARTUR PRĘDKI

METODA DEA+ W WERSJI JEDNOPRODUKTOWEJ[1]

1. WSTĘP

DEA+ to dwustopniowa procedura estymacji punktowej funkcji produkcji (transformacji) oraz miernika efektywności technicznej danej jednostki produkcyjnej w ramach pewnego, semiparametrycznego modelu granicznego. Zaproponowana została w pracy Gstach (1998) i nie zyskała szerszej popularności. Jest to jednak pierwsza chronologicznie metoda, w której metodę DEA (z ang. *Data Envelopment Analysis*) wiąże się ze złożonym składnikiem losowym. Konstrukcja odpowiedniego modelu i samej metody jest wzorowana na metodzie SFA (z ang. *Stochastic Frontier Analysis*) – zob. np. Kumbhakar, Lovell (2000). Jest to więc jeden z pomostów łączących DEA z metodami analizy procesu produkcyjnego opartymi na modelach parametrycznych. Ponadto można ją uznać za poprzedniczkę obecnie powszechnie używanej metody StoNED (zob. Kuosmanen, Kortelainen, 2012) lub opracowanie w języku polskim Prędki (2012).

Wkład własny autora polega po pierwsze na uporządkowaniu i opisaniu założeń odpowiedniego modelu semiparametrycznego będącego podstawą metodologiczną, w celu zwiększenia czytelności i jasności wywodu. Niektóre z założeń nie występują *explicite* w pracy źródłowej, lecz wynikają z kontekstu i szczątkowych wzmianek. Z kolei inne przedstawione są tam w nieco zmienionej formie. Po drugie istotne są uwagi krytyczne zawarte w artykule, które mogą rzucić pewne światło na powody niskiej popularności tej metody. Oryginalny jest również przykład empiryczny zastosowania, w którym przyjęto inne rozkłady składowych złożonego czynnika losowego, niż w pracy źródłowej.

2. MODEL SEMIPARAMETRYCZNY

Rozpoczniemy od zdefiniowania wspomnianego wcześniej modelu granicznego za pomocą szeregu założeń. Jego idea została w dużej części zapożyczona z teorii modeli parametrycznych, gdzie jest obecna od końca lat 70-tych – zob. prace źródłowe Aigner, Lovell, Schmidt (1977) oraz Meeusen, Van den Broeck (1977). Pierw-

[1] Praca wykonana w ramach Badań Statutowych finansowanych przez Wydział Zarządzania Uniwersytetu Ekonomicznego w Krakowie.

sze dwa założenia są analogiczne jak w tzw. jednoproduktowym modelu Bankera (zob. Banker, 1993).

Założenie 1 Jednostki produkcyjne wytwarzają *jeden* rodzaj produktu z *m* rodzajów nakładów i posługują się tą samą technologią reprezentowaną przez niemalejącą, wklęsłą funkcję produkcji g: $X \rightarrow R$, gdzie X jest wypukłym i zwartym podzbiorem R_{0+}^m.

Założenie 2 Dane są ilości użytych nakładów i wytworzonych produktów dla *n* producentów w postaci próby $\mathcal{X}_n = ((\mathbf{x}_j, y_j) \in X \times R_{0+}, j = 1, \ldots, n)$ rozumianej jako realizacja ciągu wektorów losowych o tym samym rozkładzie. Ponadto, wektor $\mathbf{x}_j$ charakteryzuje się gęstością h:

$$\forall \mathbf{x} \in \mathrm{int}X: h(\mathbf{x}) > 0,$$

gdzie IntX to tzw. wnętrze topologiczne zbioru X.

Założenie 3 wprowadza tzw. złożony składnik losowy wskazujący, że źródłem niepewności w odniesieniu do wielkości produktu jest nie tylko nieefektywność techniczna obiektów, lecz także tzw. szoki zewnętrzne. Gstach wprowadza go w postaci multiplikatywnej o charakterze ekspotencjalnym e^{w_j}, ze względu na późniejsze symulacje, przeprowadzane na wielkościach zlogarytmowanych[2].

Założenie 3 Spełnione jest równanie:

$$\forall j = 1, \ldots, n: y_j = g(\mathbf{x}_j)\, e^{w_j}, \tag{1}$$

gdzie $w_j = v_j - u_j$ (tzw. złożony składnik losowy).

Kolejne założenia opisują własności złożonego składnika losowego oraz jego składowych v_j i u_j.

Założenie 4 Zmienne losowe w_j tworzą ciąg niezależnych zmiennych losowych o tym samym rozkładzie.

W razie potrzeby będziemy tu używać skrótu i.i.d. (ang. *independent and identically distributed*).

Założenie 5 Składniki „białoszumowe" v_j mają identyczne rozkłady parametryczne zadane symetryczną gęstością f_V będącą funkcją klasy C^1 na nośniku $(-v_{max}, v_{max})$ zależną od wektora parametrów (θ_V, v_{max}).

W części empirycznej pracy źródłowej przyjęto tu symetrycznie ucięty rozkład beta. Natomiast autor niniejszego opracowania wykorzystał w ilustracji symetrycznie ucięty rozkład normalny.

Założenie to implikuje w szczególności, iż $E(v_j) = 0$. Przy czym w kolejnej pracy Gstach (1999) unika się założenia o symetryczności gęstości, wprowadza się tylko

---

[2] W symulacjach przyjmuje on dwa warianty „rzeczywistej" funkcji produkcji: Translog oraz funkcję kawałkami Cobba-Douglasa – szczegóły Gstach (1998, s. 168). Drugi z wariantów, po zlogarytmowaniu, jest więc funkcją kawałkami liniową, podobną do technologii płatami liniowej w DEA.

ograniczenie od góry nośnika przez v_{max}. Jednocześnie jednak niezależnie zakłada się kluczową własność $E(v_j) = 0$, która tu jest implikacją założenia 5. Zdaniem autora zmiana ta jest podyktowana problematycznością dolnego ograniczenia nośnika przez $-v_{max}$, przy wyprowadzaniu gęstości łącznej złożonego składnika losowego. Powrócimy do tego tematu w części czwartej pracy.

Założenie 6 Składniki u_j związane z nieefektywnością mają identyczne rozkłady parametryczne zadane gęstością f_U będącą funkcją klasy C^1 na nośniku R_+ zależną od parametrów θ_U.

W pracy źródłowej przyjęto w symulacjach rozkład półnormalny, natomiast autor niniejszego artykułu posłużył się w ilustracji rozkładem wykładniczym. W razie „bogatszej” parametryzacji odpowiednich rozkładów wielkości θ_U, θ_V mogą być wektorami parametrów.

Autor metody nie zakłada *explicite* niezależności ciągu zmiennych losowych v_j oraz niezależności ciągu zmiennych losowych u_j. W pracy Gstach (1998, przypis 5) zaznacza się jedynie, iż założenie 4 można osłabić przyjmując, że tylko zmienne losowe v_j tworzą ciąg i.i.d..

W kwestii niezależności u_j, v_j oraz wektora $\mathbf{x}_j$ wspomina się, iż nakłady są wielkościami egzogenicznymi, zaś nieefektywność ma wpływ jedynie na wielkość produktu. Jednak, przy wyprowadzaniu postaci tzw. funkcji wiarygodności, korzysta się z niezależności odpowiednich składowych. Przyjmiemy więc jeszcze jedno założenie.

Założenie 7 Składowe u_j, v_j są niezależne od siebie i od wektora $\mathbf{x}_j$.

3. METODA DEA+

Przejdźmy teraz do samej metody DEA+ bazującej na przedstawionym modelu. Na początek wprowadźmy pewną definicję.

Definicja 1 Funkcję $\tilde{g}(\mathbf{x}) = g(\mathbf{x})e^{v_{max}}$ nazywamy pseudo-granicą produkcyjną, zaś składnik losowy $\tilde{w}_j = w_j - v_{max}$ nazywamy pseudo-efektywnością obiektu j-tego.

Zauważmy, że:

$$y_j = g(\mathbf{x}_j)e^{w_j} = g(\mathbf{x}_j)e^{w_j - v_{max} + v_{max}} = \tilde{g}(\mathbf{x}_j)e^{\tilde{w}_j}, \tag{2}$$

Ponadto:

$$\tilde{w}_j = w_j - v_{max} = (v_j - v_{max}) - u_j \le 0, \tag{3}$$

Kierunek nierówności wynika z postaci nośników gęstości przyjętych w założeniach 5 i 6.

Oznacza to, że na obserwowalną wielkość produktu można patrzeć dwojako. W rzeczywistości obserwuje się maksymalną wielkość produkcji zaburzoną przez oba rodzaje szoków (zewnętrzny i nieefektywność). Z drugiej strony, można przyjąć, że jest to maksymalna wartość pseudogranicy zaburzona przez szok, który traktuje się jak specyficzną nieefektywność związaną z ową pseudogranicą. Tę drugą koncepcję wykorzystano w I etapie metody DEA+. Wprost z definicji wynika, iż kształt pseudogranicy jest identyczny jak prawdziwej funkcji produkcji g (jest ona tylko przesunięta o wielkość $e^{v_{max}}$). Podobnie sprawa ma się z rozkładem odchylenia $\tilde{w}_j$. Na mocy założenia 4 oraz wprost z definicji 1, pseudoefektywności tworzą ciąg i.i.d., a ich rozkład jest rozkładem złożonego składnika losowego w_j przesuniętym w dół o wielkość v_{max}.

Jak wspomniano na początku, DEA+ jest procedurą dwuetapową. W etapie I stosuje się zwykłą metodę DEA do estymacji pseudogranicy i pseudoefektywności dla obiektu j-tego, w sposób analogiczny jak w jednoproduktowym modelu Bankera (por. Banker, 1993). Oznacza to, iż:

$$\hat{\tilde{g}}(\mathbf{x}_j) = \max\{y: \exists\lambda_j \geq 0: \mathbf{x}_o \geq \sum_{j=1}^{n} \lambda_{jo}\mathbf{x}_j,\ y \leq \sum_{j=1}^{n} \lambda_{jo}\mathbf{y}_j,\ \sum_{j=1}^{n} \lambda_{jo} = 1\}. \tag{4}$$

Zaś ze względu na multiplikatywną postać złożonego składnika losowego:

$$\hat{\tilde{w}}_j = \ln(y_j) - \ln(\hat{\tilde{g}}(\mathbf{x}_j)). \tag{5}$$

Sensowność etapu I legitymizuje następujące twierdzenie.

Twierdzenie 1 Przy wprowadzonych założeniach 1–7:

– $\hat{\tilde{g}}(\mathbf{x}_j)$ jest zgodnym estymatorem $\tilde{g}(\mathbf{x}_j)$, dla $\mathbf{x}_j \in \mathrm{int}X$;

– asymptotyczny rozkład odchylenia $\hat{\tilde{w}}_j$ jest identyczny jak składnika losowego $\tilde{w}_j$.

Dowód: Teza wynika z twierdzeń 5 i 6 zawartych w pracy Banker (1993) zastosowanych w odniesieniu do wielkości $\tilde{g}(\mathbf{x}_j)$ oraz $\tilde{w}_j$ (wszystkie założenia tych twierdzeń są tu spełnione).

W etapie II metody DEA+ oblicza się za pomocą MNW oceny parametrów $\theta = (\theta_U, \theta_V, v_{max})$, opierając się na ocenach pseudoefektywności $\hat{\tilde{w}}_j$, obliczonych w etapie I. Oznacza to, iż:

$$(\hat{\theta}_U, \hat{\theta}_V, \hat{v}_{max}) = \arg\max_\theta \ln\left[\prod_{j\in J} f_{\tilde{w}}\left(\hat{\tilde{w}}_j|\theta\right)\right], \tag{6}$$

gdzie

$$f_{\tilde{w}}\left(\hat{\tilde{w}}\right) = \int_{\hat{\tilde{w}}}^{0} f_V\left(v + v_{max}\right) f_U\left(v - \hat{\tilde{w}}\right) dv, \tag{7}$$

oraz

$$J = \{j: \hat{\tilde{w}}_j < 0\}. \tag{8}$$

Wprowadzenie zbioru J oznacza, że w estymacji bierze się pod uwagę wyłącznie tzw. obiekty pseudonieefektywne czyli takie, że $\hat{\tilde{w}}_j < 0$. Dla obiektów pseudoefektywnych, gdzie $\hat{\tilde{w}}_j = 0$, formuła (6) ulega degeneracji. Ponadto nie można wtedy wykluczyć, iż $\mathbf{x}_j \notin \mathrm{int}X$. Co z kolei oznacza, że metoda DEA+ może być niezgodna. Gstach wskazuje jednak, iż asymptotycznie frakcja obiektów poza zbiorem J jest zaniedbywalna, tzn. zachodzi:

$$\lim_{n\to\infty} \frac{\left[\sum_{j=1}^{n} \mathrm{I}_{j\notin J}(j)\right]}{n} = 0\,, \tag{9}$$

gdzie I(·) – funkcja wskaźnikowa. Tak więc w sensie asymptotycznym nie będzie miało znaczenia czy procedurę wykonuje się na wszystkich obserwacjach, czy tylko na tych pseudonieefektywnych.

Mając oceny wszystkich parametrów charakteryzujących rozkłady obu składowych złożonego składnika losowego można po pierwsze obliczyć ocenę punktową funkcji produkcji:

$$\hat{g}(\mathbf{x}_j) = \hat{\hat{g}}(\mathbf{x}_j)\, e^{-\hat{v}_{max}}\,. \tag{10}$$

Przy użyciu metod analogicznych jak w SFA można też otrzymać ocenę miernika efektywności j-tego obiektu (zob. np. Kumbhakar, Lovell, 2000). Przy czym autor metody jedynie wspomina o tej możliwości w pracy Gstach (1999) nie realizując jej[3]. Warto zaznaczyć, iż uzyskując w szczególności oceny wariancji obu składowych czynnika losowego, nie zachodzi konieczność stosowania metody momentów czy pseudowiarygodności. Pierwsza z nich choć prosta nie jest pozbawiona wad i ograniczeń – zob. Kumbhakar, Lovell (2000). Druga natomiast jest dość żmudna od strony obliczeniowej – zob. np. Kuosmanen, Kortelainen (2012).

Według Gstacha (1998) zgodność estymatora wykorzystywanego w etapie I jest uwarunkowana wprowadzeniem ograniczoności nośnika „szumu". Jednostronny składnik losowy $\tilde{w}_j$, szacowany za pomocą estymatora DEA o analogicznej własności, jest bowiem konstruowany za pomocą parametru v_{max} – zob. wzór (3). Autor niniejszej pracy chciałby jednak zwrócić uwagę na inną rolę parametru v_{max}. Stanowi on niezbędny składnik korekty wyjściowej granicy produkcyjnej. Jest to więc postępowanie podobne jak w metodach SMNK czy zmodyfikowanej MNK, gdzie wyjściową ocenę granicy produkcyjnej w punktach danych również korygujemy tyle, że odpowiednio o największą resztę lub o charakterystykę $E(u_j)$.

[3] Oblicza się jedynie charakterystykę $E(U|\hat{\theta}_U)$ zwaną średnią nieefektywnością, zaś głównym celem symulacji jest porównanie metod SFA i DEA+.

W części trzeciej pracy źródłowej Gstach (1998) autor stara się dowieść zgodność całej procedury DEA+. W szczególności, opierając się na twierdzeniu 1 (zgodność etapu I) oraz twierdzeniu z pracy Bierens (1994), autor metody dowodzi twierdzenie o zgodności oceny granicy produkcyjnej w punktach danych uzyskanej w wyniku zastosowania DEA+.

Twierdzenie 2 Przy wprowadzonych założeniach 1–7 metoda DEA+ dostarcza zgodnego estymatora $\hat{g}(\mathbf{x}_j)$ wartości $g(\mathbf{x}_j)$, dla $\mathbf{x}_j \in \mathrm{int}X$.

Dowód: Gstach (1998, s. 165–167).

4. UWAGI KRYTYCZNE

Pierwsza uwaga dotyczy nieprawidłowości występującej w przeprowadzonej procedurze MNW, kwestionuje więc tym samym zgodność metody DEA+. W zapisie funkcji wiarygodności we wzorze (6) pojawia się iloczyn gęstości zmiennych losowych $\hat{\tilde{w}}_j$ mimo, że nie muszą one być niezależne, na co zresztą zwraca uwagę sam autor. Powołując się na pracę (Banker 1993) twierdzi on, iż asymptotycznie zmienne te „stają się coraz bardziej niezależne", podobnie jak oceny odchyleń w jednoproduktowym modelu Bankera. Jednak argumentacja użyta w pracy Bankera wydaje się mocno wątpliwa. Po pierwsze jest mocno lakoniczna i składa się z kilku słów odnośnika w przypisie 7 do pracy Tate, Brown (1970), w której jakoby prezentowany jest podobny punkt widzenia. Po drugie wspomniana praca z 1970 roku dotyczy własności testu Q Cochrana, badanych za pomocą odpowiednich eksperymentów symulacyjnych. Autor niniejszej pracy nie dostrzega jej związku z kwestią niezależności ocen odchyleń przy rosnącej liczbie obserwacji, a już tym bardziej z wpływem tego zjawiska na własności statystyczne MNW. Ponadto procedura DEA+ używana jest w praktycznych zastosowaniach na próbach skończonych, gdzie niezależność nie jest zagwarantowana, a własności małopróbkowe procedury nie są znane. Autor metody wykonał jedynie badania symulacyjne w celu ich poznania na wybranych przykładach (zob. Gstach 1998).

Druga uwaga dotyczy wzoru na gęstość łączną $\tilde{W}$ podanego w pracy Gstach (1998, s. 165) używanego w II etapie metody dla reszt $\hat{\tilde{w}}_j$ – por. wzór (7):

$$\mathrm{f}_{\tilde{W}}(\tilde{w}) = \int_{\tilde{w}}^{0} \mathrm{f}_{V}(v + v_{\max}) \mathrm{f}_{U}(v - \tilde{w}) dv \,. \tag{11}$$

Sam wzór w ogólności jest prawdziwy i wynika to z faktu, że zmienna losowa $\tilde{W}$ jest sumą niezależnych zmiennych losowych –U oraz V $-v_{\max}$. Jej gęstość jest wtedy splotem gęstości odpowiednich zmiennych losowych. Korzystając z własności:

$$\mathrm{f}_{-X}(x) = \mathrm{f}_X(-x),\ \mathrm{f}_{X-C}(x) = \mathrm{f}_X(x + C), \tag{12}$$

gdzie X – zmienna losowa, C – stała, można bez problemu dowieść prawdziwości powyższego wzoru.

Wątpliwości autora niniejszej pracy budzi natomiast zasadność granic całkowania przyjętych we wzorze. Z postaci nośników zmiennych V i U zawartych w założeniach 5 i 6 wynika, iż:

$$F_V(v + v_{max}) > 0, \text{ dla } |v + v_{max}| < v_{max}, \tag{13}$$

$$f_U(v - \tilde{w}) > 0, \text{ dla } v - \tilde{w} > 0.$$

Ponieważ całkuje się po v, powyższe nośniki można przekształcić do postaci:

$$-2v_{max} < v < 0,\ v > \tilde{w}. \tag{14}$$

Korzystając z faktu, iż $\tilde{w} \le 0$, przedział $(\tilde{w}, 0)$ rzeczywiście stanowiłby część wspólną nierówności (14), gdyby nie wymagana nierówność $-2v_{max} < v$. Okazuje się, iż z przyczyn technicznych ma wtedy znaczenie jaka relacja łączy wielkości $\tilde{w}$ oraz $-2v_{max}$. Jeśli $\tilde{w} > -2v_{max}$ to wspomniana nierówność jest spełniona i wtedy rzeczywiście granice całkowania przyjęto prawidłowo. Jeśli jednak relacja jest przeciwna to powinno się całkować po przedziale $(-2v_{max}, 0)$, gdzie obie gęstości w iloczynie pod całką są dodatnie. Warto jednak zaznaczyć, że przyjmując w drugiej sytuacji granice całkowania po przedziale $(\tilde{w}, 0)$ wzór (11) w dalszym ciągu jest prawdziwy z tym, że odpowiednia całka na przedziale $(\tilde{w}, -2v_{max})$ wynosi zero.

Niestety pozostawienie wyjściowych granic całkowania $(\tilde{w}, 0)$ w przypadku gdy $\tilde{w} < -2v_{max}$ powoduje brak zbieżności procedury[4] realizującej MNW na resztach $\hat{\tilde{w}}$, używanej przez dodatek Solver w programie Microsoft Excel. Przy różnych punktach startowych i odpowiednich ograniczeniach nałożonych na parametry optymalna wartość v_{max} przyjmuje niedopuszczalną zerową wartość, bądź zatrzymuje się na narzuconym ograniczeniu dolnym. Możliwe jest oczywiście, iż użyto niewłaściwej metody, bądź nieprawidłowo ją zaimplementowano. Autor niniejszej pracy sądzi jednak, że należy po prostu zmodyfikować wzór (11) do postaci:

$$f_{\tilde{w}}(\tilde{w}) = \begin{cases} \int_{\tilde{w}}^{0} f_V(v + v_{max}) f_U(v - \tilde{w})\,dv, & \tilde{w} > -2v_{max} \\ \int_{-2v_{max}}^{0} f_V(v + v_{max}) f_U(v - \tilde{w})\,dv, & \tilde{w} < -2v_{max} \end{cases}. \tag{15}$$

[4] Jest to tzw. metoda gradientu sprzężonego ozn. angielskojęzycznym skrótem GRG (*Generalized Reduced Gradient*).

Funkcja wiarygodności zmienia bowiem swą postać w zależności od tego jaka relacja łączy reszty $\hat{\hat{w}}$ i nieznany parametr – $2v_{max}$, który podlega optymalizacji.

Prawdopodobnie właśnie z tego powodu w pracy Gstach (1999) rezygnuje się z dolnego ograniczenia zmiennej V i jednocześnie zakłada się pożądaną dla „szumu" własność E(V) = 0. Brak jest jednak konkretnych propozycji użytecznych rozkładów składnika losowego v o nośniku nieograniczonym od dołu, który spełniałby te warunki.

5. PRZYKŁAD EMPIRYCZNY

Wykorzystano zestaw danych zaczerpnięty z pracy Osiewalski, Wróbel-Rotter (2002), używany już wielokrotnie przez autora w jego poprzednich pracach. Są to dane rzeczywiste z roku 1995 dotyczące 32 polskich elektrowni i elektrociepłowni, których efektywność techniczną będziemy analizować. Za nakłady przyjęto:
– kapitał (wartość brutto środków trwałych liczona w zł.);
– pracę (liczba pracowników);
– energię wsadu (liczoną w TJ).

Jedynym produktem działalności jednostek jest wytworzona energia (liczona w TJ[5]).

Zrealizowano pierwszy etap metody DEA+ za pomocą wzoru (4), obliczając na wyjściowych danych standardowy estymator DEA $\hat{\hat{g}}(\mathbf{x}_j)$, j = 1,…,*n*. Następnie korzystając ze wzoru (5) uzyskano reszty $\hat{\hat{w}}_j$, j = 1,…,*n*. Przypomnijmy, że są to odpowiednio oceny pseudogranicy produkcyjnej oraz pseudoefektywności w punktach danych. Rezultaty przedstawiono w tabeli 1.

W dalszej kolejności przystąpiono do realizacji etapu II. Zgodnie z wcześniejszymi uwagami do założeń 5 i 6 przyjęto przykładowo rozkład normalny $N(0, \sigma_v^2)$ ucięty do przedziału $(-v_{max}, v_{max})$ dla „szumu" v oraz rozkład wykładniczy $Exp(\lambda)$ dla składowej u związanej z nieefektywnością. Zrealizowano wzór (15) na gęstość łączną $\tilde{W}$ będący zmodyfikowaną wersją wzoru (11), zgodnie ze uwagami zawartymi w części czwartej niniejszego opracowania. Po szeregu żmudnych, lecz prostych przekształceń uzyskano ostatecznie następującą postać tej gęstości:

$$f_{\tilde{w}}(\tilde{w}) = \frac{\lambda \exp\left\{\frac{-1}{2}\left[\lambda^2\sigma_v^2 - 2\lambda\left(v_{max} + \tilde{w}\right)\right]\right\}\left[\Phi\left(\frac{v_{max} + \lambda\sigma_v^2}{\sigma_v}\right) - \Phi\left(\frac{a + \lambda\sigma_v^2}{\sigma_v}\right)\right]}{2\Phi\left(\frac{v_{max}}{\sigma_v}\right) - 1}, \quad (16)$$

[5] 1GWh=3,6TJ (teradżul).

Tabela 1

Wyniki uzyskane w etapie I metody DEA+.

Lp.	y_j	$\hat{\tilde{g}}(\mathbf{x}_j)$	$\ln y_j$	$\ln\hat{\tilde{g}}(\mathbf{x}_j)$	$\hat{\tilde{w}}_j$
1	100749	100749	11,52039	11,52039	0
2	54677	54677	10,9092	10,9092	0
3	51499,5	51499,5	10,84933	10,84933	0
4	34356,7	34356,7	10,44455	10,44455	0
5	33991,7	46883,28	10,43387	10,75542	-0,32154
6	33731,4	42135,09	10,42618	10,64864	-0,22245
7	32164,4	41151,15	10,37862	10,62501	-0,24639
8	29404,3	29404,3	10,2889	10,2889	0
9	28615,7	29431,12	10,26171	10,28981	-0,0281
10	28536,7	38289,02	10,25895	10,55292	-0,29397
11	20160,4	20160,4	9,911476	9,911476	0
12	19731,6	27213,16	9,889977	10,21146	-0,32148
13	17874,2	25937,79	9,791114	10,16346	-0,37234
14	16620,7	17007,94	9,718404	9,741436	-0,02303
15	16428,7	16428,7	9,706785	9,706785	0
16	15318,7	18621,02	9,63683	9,832046	-0,19522
17	14641,1	17974,8	9,591588	9,796726	-0,20514
18	12752,9	12752,9	9,453514	9,453514	0
19	10185,1	10527,49	9,228681	9,261745	-0,03306
20	10119,2	10593,52	9,22219	9,267998	-0,04581
21	9406,5	13667,05	9,149156	9,522743	-0,37359
22	8964	15231,13	9,100972	9,631097	-0,53013
23	7922,1	7922,1	8,977412	8,977412	0
24	6530	7890,613	8,784162	8,973429	-0,18927
25	4900,7	6260,078	8,497133	8,741948	-0,24481
26	4533,9	6487,502	8,419338	8,777633	-0,3583
27	4472,5	6130,73	8,405703	8,721069	-0,31537
28	4424,7	5164,437	8,394958	8,549551	-0,15459
29	3926,6	3926,6	8,275529	8,275529	0
30	3320,9	3320,9	8,107991	8,107991	0
31	3153,1	3153,1	8,056141	8,056141	0
32	1939	1939	7,569928	7,569928	0

Źródło: opracowanie własne.

gdzie $\Phi(\cdot)$ dystrybuanta rozkładu normalnego standaryzowanego oraz:

$$a = \begin{cases} \hat{\tilde{w}}_j + v_{\max}, \hat{\tilde{w}}_j > -2v_{\max} \\ -v_{\max}, \hat{\tilde{w}}_j < -2v_{\max} \end{cases}. \tag{17}$$

W dalszej kolejności wyprowadzono wzór na zlogarytmowaną funkcję wiarygodności opierając się na resztach $\hat{\tilde{w}}_{\mathrm{j}}$ jednostek pseudonieefektywnych.

$$\ln\left[\prod_{\mathrm{j}\in \mathrm{J}} \mathrm{f}_{\tilde{w}}\left(\hat{\tilde{w}}_{\mathrm{j}}|\theta\right)\right] = \tag{18}$$

$$\sum_{j\in J}\left\{\ln\lambda - \frac{1}{2}\left[\lambda^2\sigma_v^2 - 2\lambda\left(v_{\max} + \hat{\tilde{w}}\right)\right] + \ln\left[\Phi\left(\frac{v_{\max} + \lambda\sigma_v^2}{\sigma_v}\right) - \Phi\left(\frac{a + \lambda\sigma_v^2}{\sigma_v}\right)\right] - \ln\left[2\Phi\left(\frac{v_{\max}}{\sigma_v}\right) - 1\right]\right\}.$$

W celu uzyskania zbieżności procedury realizującej MNW narzucono następujące restrykcje na parametry:

$$\lambda,\ v_{\max},\ \sigma_v \geq 10^{-4};\ v_{\max}/\sigma_v \geq 10^{-4};\ \frac{v_{\max} + \lambda\sigma_v^2}{\sigma_v} \in [-10, 10];\ \hat{\tilde{w}}_{\mathrm{j}}/\sigma_v \leq -10^{-4},\ \mathrm{j} \in \mathrm{J}. \tag{19}$$

Ograniczenia te zapewniają „poprawne" zachowanie wyrażeń będących argumentami dystrybuant zawartych we wzorze (18) na logarytm funkcji wiarygodności w trakcie trwania algorytmu. Za punkt startowy przyjęto ostatecznie wartości ($\sigma_v;\ \lambda;\ v_{\max}$) = (0,01; 0,01; 0,011515828). Startowa wartość $v_{\max}$ spełnia zależność $\max\limits_{\mathrm{j}\in \mathrm{J}} \hat{\tilde{w}}_{\mathrm{j}} = -2v_{\max}$. Jest więc ściśle związana z kluczową relacją między resztami a odpowiednim parametrem. Istotne znaczenie dla uzyskania zbieżności MNW ma tu użycie największej reszty wśród obiektów nieefektywnych[6].

Ostatecznie uzyskano następujące oceny parametrów:

$$\left(\hat{\sigma}_v, \hat{\lambda}, \hat{v}_{\max}\right) = (0{,}002014603;\ \ 4{,}642198937;\ 0{,}020127193). \tag{20}$$

[6] Nasuwa się skojarzenie z metodą SMNK, gdzie największa reszta również odgrywa ważną rolę.

Jedynie ograniczenie $\frac{v_{\max} + \lambda\sigma_v^2}{\sigma_v} \leq 10$ jest w tym punkcie spełnione słabo. Za pomocą oceny $\hat{v}_{\max}$ oszacowano rzeczywistą granicę produkcyjną w punktach danych ze wzoru (10). Następnie oceny parametrów wykorzystano do obliczenia wartości miernika efektywności dla poszczególnych obiektów wg schematu opisanego w pracy Kumbhakar i Lovell (2000, s.78-82). Miernik ów dany jest wzorem[7]:

$$\mathrm{TE}_j = \exp(-\hat{\mathrm{E}}(u_j|\tilde{w}_j)),\ j = 1,\ldots,n. \tag{21}$$

Aby zrealizować formułę (21) należy w pierwszej kolejności uzyskać postać rozkładu warunkowego $U|\tilde{w}$. Korzystając z danych gęstości zmiennych losowych U i V, po dokonaniu prostych przekształceń otrzymuje się:

$$U|\tilde{w} \sim C(\tilde{w})\cdot N^+(\mu, \sigma_v^2). \tag{22}$$

Odpowiedni rozkład warunkowy jest więc iloczynem rozkładu normalnego uciętego o średniej:

$$\mu = -\tilde{w} - v_{\max} - \sigma_v^2/\sigma_u, \tag{23}$$

i wyrażenia:

$$C(\tilde{w}) = \exp\left(\frac{\sigma_v^2}{\sigma_u^2}\right)\frac{\Phi(b)}{\Phi(c)-\Phi(b)}, \tag{24}$$

gdzie $\sigma_u = 1/\lambda$ (odchylenie standardowe U) oraz:

$$b = \begin{cases} \dfrac{(\tilde{w} + v_{\max})\sigma_u + \sigma_v^2}{\sigma_u\sigma_v}, \tilde{w} > -2v_{\max} \\ \dfrac{(-v_{\max})\sigma_u + \sigma_v^2}{\sigma_u\sigma_v}, \tilde{w} < -2v_{\max} \end{cases}, \tag{25}$$

$$c = \frac{v_{\max}\sigma_u + \sigma_v^2}{\sigma_u\sigma_v}. \tag{26}$$

[7] Jest to jedna z propozycji pomiaru efektywności przedstawiona we wspomnianej pracy źródłowej.

Przy obliczaniu wartości oczekiwanej występującej we wzorze (21) skorzystano ze wzorów zawartych w pracy Kumbhakar, Lovell (2000, s. 82), dokładając brakujące wyrażenie $C(\tilde{w})$ dane wzorem (24). W zapisie dla obserwacji j = 1,…,*n* uzyskano:

$$E\,(u_j|\,\tilde{w}_j) = C(\tilde{w}_j)[\mu_j + \sigma_v \frac{\varphi(-\mu_j/\sigma_v)}{\Phi(\mu_j/\sigma_v)}], \tag{27}$$

gdzie $\mu_j = -\tilde{w}_j - v_{max} - \sigma_v^2/\sigma_u$ oraz $\varphi(\cdot)$ gęstość rozkładu normalnego standaryzowanego. Zrealizowano wzór (27) na resztach $\hat{\tilde{w}}_j$ oraz ocenach parametrów uzyskanych za pomocą metody DEA+ i uzyskano ocenę wartości oczekiwanej występującą we wzorze (21), a tym samym otrzymano miernik TE_j. Wyniki dla badanych jednostek produkcyjnych[8] zestawiono w tabeli 2.

Widoczne jest, iż metoda ulega degeneracji i nie może być stosowana dla obiektów pseudoefektywnych, o czym wspominano wcześniej. Na koniec podkreślmy, iż przykład ten ze względu na małą liczebność próby ma jedynie ilustrować działanie metody. Uzyskane rezultaty należy traktować z dużą ostrożnością ze względu na asymptotyczne własności DEA+.

Tabela 2

Oszacowanie granicy produkcyjnej i miary efektywności w punktach danych.

j	$\hat{\tilde{g}}(\mathbf{x}_j)$	TE_j
1	98741,48	x
2	53587,51	x
3	50473,32	x
4	33672,11	x
5	45949,09	0,739779
6	41295,5	0,816842
7	40331,17	0,797519
8	28818,39	x
9	28844,68	0,99208

[8] Dla obiektów pseudoefektywnych wartości miary TE_j nie da się obliczyć ze względu na to, iż w mianowniku wyrażenia $C(\hat{\tilde{w}}_j)$ pojawia się zero.

j	$\hat{\hat{g}}(\mathbf{x}_j)$	TE_j
10	37526,07	0,76046
11	19758,68	x
12	26670,91	0,739827
13	25420,96	0,703137
14	16669,04	0,996808
15	16101,34	x
16	18249,98	0,839395
17	17616,63	0,831108
18	12498,79	x
19	10317,71	0,987165
20	10382,44	0,974664
21	13394,72	0,702263
22	14927,64	0,600502
23	7764,244	x
24	7733,385	0,844404
25	6135,34	0,798777
26	6358,232	0,713084
27	6008,569	0,744363
28	5061,53	0,874196
29	3848,359	x
30	3254,728	x
31	3090,271	x
32	1900,363	x

Źródło: opracowanie własne.

6. ZAKOŃCZENIE

W niniejszej pracy przedstawiono wersję jednoproduktową metody DEA+ tzn. przyjęto, iż badana grupa jednostek produkcyjnych wytwarza tylko jeden produkt. Taki właśnie wariant przedstawiony jest w źródłowej pracy Gstach (1998). Docelowo jednak, zamierzeniem autora będzie konstrukcja i wykorzystanie wersji wieloproduktowej metody DEA+, zgodnie z dość ogólnymi wskazówkami zawartymi w pracy Gstach (1999). Jest to konieczne ze względu na to, iż DEA stosowana jest zwykle w takiej właśnie sytuacji.

Uniwersytet Ekonomiczny w Krakowie

LITERATURA

Aigner D., Lovell C. A. K., Schmidt P., (1977), Formulation and Estimation of Stochastic Frontier Models, *Journal of Econometrics*, 6, 21–37.

Banker R. D., (1993), Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation, *Management Science*, 39 (10), 1265–1273.

Bierens H. J., (1994), *Topics in Advanced Econometrics*, Cambridge University Press, Cambridge.

Gstach D., (1998), Another Approach to Data Envelopment Analysis in Noisy Environments: DEA+, *Journal of Productivity Analysis*, 9, 161–176.

Gstach D., (1999), Technical Efficiency in Noisy Multi-Output Settings, *CEJOR* 7, 93–110.

Kumbhakar S. C., Lovell C. A. K., (2000), *Stochastic Frontier Analysis*, Cambridge University Press, Cambridge.

Kuosmanen T., Kortelainen M., (2012), Stochastic Non-Smooth Envelopment of Data: Semi-Parametric Frontier Estimation Subject to Shape Constraints, *Journal of Productivity Analysis*, 38, 11–28.

Meeusen W., Van den Broeck J., (1977), Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error, *International Economic Review*, 18 (2), 435–444.

Osiewalski J., Wróbel-Rotter R., (2002), Bayesowski model efektów losowych w analizie efektywności kosztowej (na przykładzie elektrowni i elektrociepłowni polskich), *Przegląd Statystyczny*, 50 (2), 47–68.

Prędki A., (2012), Wybrane metody estymacji w semiparametrycznym modelu granicznym, *Przegląd Statystyczny*, 59 (3), 215–232.

Tate M. W., Brown S. M., (1970), Note on the Cochran Q test, *Journal of the American Statistical Association*, 65 (329), 155–160.

METODA DEA+ W WERSJI JEDNOPRODUKTOWEJ

Streszczenie

W artykule opisano DEA+, która jest jedną z metod służących do estymacji funkcji produkcji oraz miar efektywności technicznej. W pracy rozważono przypadek jednoproduktowy. Przedstawiono semiparametryczny model graniczny, będący podstawą teoretyczną metody oraz sam jej przebieg, wraz z uwagami krytycznymi dotyczącymi prawidłowości funkcjonowania procedury. Całość kończy przykład empiryczny ilustrujący zastosowanie metody, przy wybranych założeniach modelowych w odniesieniu do rozkładów składowych złożonego czynnika losowego.

Słowa kluczowe: DEA+, semiparametryczny model graniczny, funkcja produkcji, efektywność techniczna

SINGLE-PRODUCT VERSION OF THE DEA+ METHOD

A b s t r a c t

In the article we present the DEA+ method as a tool for estimation of production function and technical efficiency measures. We restrict the scope of the study only to the single-product case. Once the underlying, semiparametric frontier model is discussed, we proceed with demonstrating the very algorithm of DEA+, and provide some critique of its validity. Finally, the method is illustrated with an empirical analysis under certain choices of distributions for each of the random variables constituting the composed error.

Keywords: DEA+, semiparametric frontier model, production function, technical efficiency