PRZEGLĄD STATYSTYCZNY
R. LVII – ZESZYT 4 – 2010

ARTUR PRĘDKI

ESTYMACJA ZBIORU MOŻLIWOŚCI PRODUKCYJNYCH W RAMACH FORMALNEGO MODELU STATYSTYCZNEGO[1]

1. WSTĘP

Niniejsze opracowanie stanowi kontynuację badań opisanych w pracach [14], [16], których celem było przedstawienie możliwych rozszerzeń metod FDH i DEA[2] w ujęciu stochastycznym. Jednym z przedmiotów analizy jest tu zbiór możliwości produkcyjnych, który wyraża technologię n jednostek produkcyjnych, produkujących z p nakładów q produktów. W zależności od przyjętych o nim założeń, proponowano różne jego postacie o charakterze deterministycznym (zob. np. [5], [15]). Wspomniane założenia miały charakter nieparametryczny, co wyrażało się brakiem analitycznej specyfikacji granicy produkcyjnej[3]. Powodem była niechęć do arbitralnego określania jej parametrycznej postaci oraz poważne trudności w modelowaniu dla przypadku wielu produktów.

Błędy w zebranych danych, niepełny zbiór danych, nieuwzględnienie ważnych informacji o badanym procesie produkcyjnym czy niejednorodność technologiczna obiektów, stanowią potencjalne źródła niepewności dotyczącej rzeczywistej postaci zbioru możliwości produkcyjnych. Dopiero w latach 90. pojawiły się prace [1], [10], [11], w których przyjęto, że jest on **nieznany** i jako taki może być przedmiotem estymacji. Aby jednak stał się on przedmiotem wnioskowania statystycznego na podstawie próby należało zdefiniować odpowiedni model statystyczny. Ze względu na nieparametryczny charakter obu metod zrezygnowano z modeli parametrycznych, które wymagają jawnej specyfikacji analitycznej postaci granicy produkcyjnej. Wykorzystano natomiast modele nieparametryczne, używane już od lat 60. ubiegłego wieku do estymacji zbiorów i funkcji (zob. [3], [17], [18]).

Celem opracowania jest przedstawienie propozycji modelowania zbioru produkcyjnego na podstawie prac [10], [11], [12]. Odbędzie się to według następującego schematu. Na początek (w części drugiej) zostanie wprowadzony ogólny model statystyczny. W części tej przedstawimy też kryterium optymalności estymatorów oparte na pojęciu asymptotycznego ryzyka minimaksowego. W części trzeciej zostanie skonstruowany wariant szczegółowy modelu, służący jako podstawa estymacji zbioru możliwo-

---

[1] Praca wykonana w ramach badań statutowych finansowanych przez Uniwersytet Ekonomiczny w Krakowie.

[2] FDH z ang. *Free Disposal Hull*, DEA z ang. *Data Envelopment Analysis*.

[3] Zwanej funkcją produkcji (jeden produkt) lub transformatą produkcji (wiele produktów).

ści produkcyjnych. Przedstawimy tam też definicje odpowiednich estymatorów. Część czwarta pracy zawiera ilustrację i własności statystyczne wprowadzonych estymatorów. W zakończeniu wyniki z niniejszej pracy zostaną krótko zestawione z innymi rezultatami w zakresie estymacji zbiorów o podobnym charakterze. Wskazano również inne potencjalne zastosowania ogólnego modelu statystycznego wprowadzonego w części drugiej pracy.

2. OGÓLNY MODEL STATYSTYCZNY

Za pomocą kolejnych założeń określimy ogólny model statystyczny, który zostanie uszczegółowiony dla naszych potrzeb w podrozdziale trzecim. Zacznijmy od określenia przedmiotu naszego wnioskowania, czyli przestrzeni parametrów.

Założenie 1. Parę $(\Phi, d(\cdot,\cdot))$ nazywamy **przestrzenią parametrów** z pseudometryką $d(\cdot,\cdot)$, $\varphi \in \Phi$ jest więc **nieznanym parametrem** podlegającym estymacji.

UWAGA: W praktyce będziemy często wyodrębniać podzbiór Σ przestrzeni parametrów wyrażający pewne dodatkowe informacje o parametrze będącym przedmiotem estymacji. Ze względu na to, że parametrem mogą być obiekty typu zbiór czy funkcja wprowadzono ogólniejszy sposób pomiaru odległości za pomocą pseudometryki[4].

Kolejne założenie definiuje pojęcie wektora obserwacji.

Założenie 2. Wektor obserwacji (próba statystyczna) $\mathcal{X}_n$ jest elementem pewnej przestrzeni mierzalnej (X_n, σ_n) zwanej **przestrzenią próby**, gdzie n – liczba obserwacji.

Założenie trzecie określa ogólnie mechanizm generowania wektora obserwacji.

Założenie 3. Dana jest rodzina rozkładów prawdopodobieństwa $\left\{P_\varphi^n, \varphi \in \Sigma \subset \Phi\right\}$ na przestrzeni próby (X_n, σ_n). Zakładamy, że jeden z jej elementów (odpowiadający rzeczywistej wartości parametru) jest odpowiedzialny za sposób generowania obserwacji.

UWAGA: Istnieje obiekt losowy $\xi_n : (X, \sigma_X, P) \to (X_n, \sigma_n)$ zwany **próbą losową**, którego rozkład jest przedmiotem zainteresowania[5]. Jego realizacja to właśnie wektor obserwacji $\mathcal{X}_n$.

Założenie czwarte zapewnia istnienie rodziny gęstości odpowiadającej, wprowadzonej w poprzednim założeniu, rodzinie rozkładów prawdopodobieństwa.

Założenie 4. Istnieje σ-skończona miara ν_n na przestrzeni (X_n, σ_n) taka, że miary P_φ^n są w stosunku do niej absolutnie ciągłe.

UWAGA: Przy tym założeniu, z twierdzenia Radona-Nikodyma wynika, że:

$$\forall \varphi \in \Sigma \subset \Phi\, \exists f_\varphi^n : (X_n, \sigma_n) \to [0, \infty) \text{ mierzalna } \forall A \in \sigma_n : P_\varphi^n(A) = \int_A f_\varphi^n \, d\nu_n .$$

W tym przypadku funkcja f_φ^n jest nazywana **gęstością rozkładu prawdopodobieństwa** P_φ^n.

[4] Spełnia ona, podobnie jak metryka, warunek symetrii i „trójkąta". Dopuszcza jednak nieidentyczność obiektów, których odległość jest zerowa.

[5] Mówi o tym odpowiednie twierdzenie (zob. np. [13], s. 31). Dla $(X_n, \sigma_n) = (R^n, \mathcal{B}(R^n))$ odwzorowanie mierzalne ξ_n nazywamy **wektorem losowym**. Dodatkowo, przy n = 1, mówimy o **zmiennej losowej**.

Wprowadzimy teraz pojęcie estymatora parametru będącego przedmiotem wnioskowania.

Definicja 2.1. Dowolne, mierzalne odwzorowanie $\hat{\varphi}_n$: $(X_n, \sigma_n) \to (\Phi, d(\cdot,\cdot))$ nazywamy **estymatorem** (parametru φ).

UWAGA: W kwestii mierzalności $\hat{\varphi}_n$ warto zaznaczyć, że pseudometryka zadaje w przestrzeni parametrów topologię. Tak więc za σ-algebrę przyjęto tu rodzinę zbiorów borelowskich[6] $\mathcal{B}(\Phi)$.

Przejdziemy teraz do problemu zdefiniowania tzw. funkcji ryzyka (z ang. *risk function*) będącej podstawą, sformułowanego w dalszym ciągu pracy, kryterium optymalności estymatorów zwanego minimaksowym. Na gruncie statystyki parametrycznej funkcja ryzyka jest zwykle postaci[7]

$$R(\hat{\varphi}_n, \varphi) = E^n_\varphi\left[w(\hat{\varphi}_n, \varphi)\right] = \int_{X_n} w(\hat{\varphi}_n, \varphi)\,dP^n_\varphi,$$

gdzie:
$w(\cdot,\cdot)$ – dowolna nieujemna funkcja zwana **funkcją straty**[8] (z ang. *loss function*);
E^n_φ – wartość oczekiwana względem rozkładu P^n_φ (zakładamy jej istnienie).

Skończona parametryzacja modelu gwarantuje m.in. istnienie naturalnej odległości euklidesowej mierzącej dokładność estymacji. Dla naszego modelu ogólnego, rolę tej odległości będzie pełnić pseudometryka $d(\cdot,\cdot)$.

Definicja 2.2. **Funkcją ryzyka** nazywamy odwzorowanie $R(\cdot,\cdot)$ określone następująco:

$$R(\hat{\varphi}_n, \varphi) = E^n_\varphi\left[w\left(d(\hat{\varphi}_n, \varphi)\right)\right].$$

Po wprowadzeniu dostępnej nam informacji o przestrzeni parametrów Φ w postaci jej podzbioru Σ, definiujemy:

Definicja 2.3. **Maksymalnym ryzykiem** estymatora $\hat{\varphi}_n$, na zbiorze Σ nazywamy liczbę:

$$r(\hat{\varphi}_n) = \sup_{\varphi \in \Sigma} R(\hat{\varphi}_n, \varphi).$$

Przedmiotem naszego zainteresowania będzie estymator, dla którego to maksymalne ryzyko będzie najmniejsze.

---

[6] Rodzina zbiorów borelowskich jest to najmniejsza σ-algebra zawierająca wszystkie zbiory otwarte w przestrzeni $(\Phi, d(\cdot,\cdot))$.

[7] Zob. np. [8], s. 16.

[8] Jest to subiektywny miernik straty, jaką ponosimy na skutek odstępstwa estymatora od nieznanej wartości parametru. Często przyjmuje się jeszcze pewne, dodatkowe własności tej funkcji (zob. np. [10], s. 478).

Definicja 2.4. Liczbę $\inf_{\hat{\varphi}_n} r(\hat{\varphi}_n)$ nazywamy **ryzykiem minimaksowym**.

O ile ryzyko minimaksowe istnieje zawsze, o tyle estymator, którego maksymalne ryzyko byłoby najmniejsze wśród wszystkich możliwych estymatorów wcale nie musi istnieć[9] (tzw. estymator **optymalny w sensie ryzyka minimaksowego**). Stąd zwykle ogranicza się klasę rozważanych estymatorów[10] lub bada tę własność asymptotycznie. W naszym przypadku przedstawimy drugie podejście, ze względu na asymptotyczny charakter rezultatów uzyskanych w modelu szczegółowym.

Definicja 2.5. Granicę $\lim_{n \to \infty} \inf_{\hat{\varphi}_n} \mathrm{r}(\hat{\varphi}_n)$ nazywamy **asymptotycznym ryzykiem minimaksowym**.

Jeśli istnieje ta granica[11] i estymator $\hat{\varphi}_n^*$ takie, że:

$$\lim_{n \to \infty} \inf_{\hat{\varphi}_n} \mathrm{r}(\hat{\varphi}_n) = \lim_{n \to \infty} \mathrm{r}\left(\hat{\varphi}_n^*\right),$$

to mówimy, iż jest on **optymalny w sensie asymptotycznego ryzyka minimaksowego**.

Wprowadźmy teraz definicje związane z kolejnymi własnościami estymatorów o charakterze asymptotycznym.

Definicja 2.6. Estymator $\hat{\varphi}_n$ nazywamy **asymptotycznie nieobciążonym**, gdy:

$$\lim_{n \to \infty} \mathrm{E}_{\varphi}^{\mathrm{n}}\left[\mathrm{d}(\hat{\varphi}_n, \varphi)\right] = 0.$$

Definicja 2.7. Estymator $\hat{\varphi}_n$ nazywamy **zgodnym** (słabo zbieżnym, zbieżnym wg rozkładu), gdy:

$$\forall \varepsilon > 0: \lim_{n \to \infty} \mathrm{P}_{\varphi}^{\mathrm{n}}\left(\mathrm{d}(\hat{\varphi}_n, \varphi) < \varepsilon\right) = \lim_{n \to \infty} P_{\xi_n}\left\{\mathcal{X}_\mathrm{n} \in \mathrm{X}_\mathrm{n}: \mathrm{d}\left(\hat{\varphi}_n(\mathcal{X}_\mathrm{n}), \varphi\right) < \varepsilon\right\} = 1.$$

Stosuje się oznaczenia: $\hat{\varphi}_n \xrightarrow{P} \varphi$ lub $\mathrm{d}(\hat{\varphi}_n, \varphi) = \mathrm{o}_\mathrm{P}(1)$. Zwróćmy uwagę, że pseudometryka $\mathrm{d}(\cdot,\cdot)$ pełni tu ponownie rolę odległości euklidesowej w „zwykłej" definicji zgodności dla modeli parametrycznych[12].

Innym warunkiem jest tzw. ograniczoność wg rozkładu (z ang. *bounded in probability*).

---

[9] Zob. np. praca [8], s. 4. Supremum oraz infimum dla zbioru liczbowego zawsze istnieje, choć nie zawsze należy do tego zbioru.

[10] Na przykład rozważa się tylko estymatory asymptotycznie normalne albo nieobciążone parametru, zob. [8], s. 4 i 17.

[11] Przykład zaimplementowania tej idei, przy pewnych dodatkowych warunkach narzuconych na funkcję straty, w pracy [10], s. 478-479.

[12] Tę koncepcję rozszerzenia definicji własności asymptotycznych można znaleźć np. w pracy [20], s. 16.

Definicja 2.8. Estymator $\hat{\varphi}_n$ nazywamy **ograniczonym wg rozkładu**, gdy[13]:

$$\forall \varepsilon > 0 \, \exists B_\varepsilon \, \exists n_\varepsilon \, \forall n \geq n_\varepsilon \, P^n_\varphi \left(d(\hat{\varphi}_n, \varphi) \leq B_\varepsilon \right) \geq 1 - \varepsilon.$$

Stosuje się oznaczenie[14]: $d(\hat{\varphi}_n, \varphi) = O_P(1)$.

Warunek zgodności jest tu najsilniejszy, o czym mówi poniższe twierdzenie[15].

Twierdzenie 2.1. Jeśli $\hat{\varphi}_n$ jest estymatorem o skończonej wartości oczekiwanej, zbieżnym wg rozkładu, to jest również ograniczony wg rozkładu oraz asymptotycznie nieobciążony.

Warto zastanowić się nad rolą powyższego, słabszego warunku w dalszej analizie. Czy jego wprowadzenie oznacza, że nie mogąc dowieść zbieżności wg rozkładu, będziemy starać się dowieść choćby ograniczoność wg rozkładu? Zdecydowanie nie, jego rola jest tu „bardziej pozytywna" i wiąże się ściśle z pojęciem tempa zbieżności (z ang. *rate of convergence*).

Definicja 2.9. **Tempem zbieżności** nazywamy dowolny, dodatni, zbieżny do zera ciąg liczb rzeczywistych ψ_n.

Pojęcie to służy do określenia „szybkości" zachodzenia różnego typu własności asymptotycznych (np. zbieżności wg rozkładu, ale również optymalności w sensie ryzyka minimaksowego).

Wprowadźmy to pojęcie do użytku w kontekście dwóch zdefiniowanych powyżej rodzajów zbieżności.

Definicja 2.10. Estymator $\hat{\varphi}_n$ jest zbieżny (ograniczony) wg prawdopodobieństwa w **tempie** ψ_n, gdy:

$$\psi_n^{-1} d(\hat{\varphi}_n, \varphi) = o_P(1) \quad \left[\psi_n^{-1} d(\hat{\varphi}_n, \varphi) \right) = O_P(1) \right].$$

Stosuje się, odpowiednio, oznaczenie: $d(\hat{\varphi}_n, \varphi) = o_P(\psi_n)$ oraz $d(\hat{\varphi}_n, \varphi) = O_P(\psi_n)$.

Zwróćmy uwagę, że ze zbieżności ciągu ψ_n^{-1} do nieskończoności, wynika odpowiednio zbieżność (ograniczoność) wg rozkładu ciągu zmiennych losowych $d(\hat{\varphi}_n, \varphi)$. Ponadto zachodzą następujące zależności.

Twierdzenie 2.2. Jeśli $d(\hat{\varphi}_n, \varphi) = O_P(\psi_n)$, to $d(\hat{\varphi}_n, \varphi) = o_P(1)$.

---

[13] Definicja za pracą [21], s. 437-438. Zapis skrócony (por. def. 2.7).

[14] Oznaczenia „O-duże" i „o-małe" stosuje się również w analizie matematycznej dla określenia szybkości zbieżności ciągów czy funkcji rzeczywistych.

[15] Zob. np. [20], s. 60-61. W pracy tej opisano również inne związki pomiędzy różnymi rodzajami zbieżności zmiennych losowych.

Dowód: Implikacja zachodzi na mocy wprowadzonych definicji oraz faktu, że:

$$\forall \varepsilon > 0 \forall B_\varepsilon > 0: B_\varepsilon \psi_n \to 0.$$

Twierdzenie 2.3. Jeśli $d(\hat{\varphi}_n, \varphi) = O_P(\psi_n)$ i $\forall n \in N: \underline{\psi}_n \geq \psi_n, \underline{\psi}_n \to 0$, to $d(\hat{\varphi}_n, \varphi) = O_P(\underline{\psi}_n)$.

Dowód: Twierdzenie wynika wprost z odpowiednich definicji oraz faktu, że:

$$\forall n \in N: \underline{\psi}_n^{-1} \leq \psi_n^{-1}.$$

Z powyższych twierdzeń wynika, iż własność ograniczoności wg prawdopodobieństwa może służyć wykazaniu zgodności innego ciągu zmiennych losowych. Ponadto, dany ciąg zmiennych losowych może mieć wiele różnych temp zbieżności. Byłoby interesującym wykazanie, że wśród nich istnieje tempo optymalne w jakimś sensie. W celu znalezienia owego kryterium optymalności wykorzystamy kryterium minimaksowe. Tym razem jednak definicje będą ukierunkowane na poszukiwanie optymalnego tempa zbieżności.

Definicja 2.11. **Funkcją ryzyka R(·,·,·) w tempie**[16] $\boldsymbol{\psi_n}$ nazywamy odwzorowanie określone następująco:

$$R(\hat{\varphi}_n, \varphi, \psi_n) = E_\varphi^n\left[w\left(\psi_n^{-1} \cdot d(\hat{\varphi}_n, \varphi)\right)\right].$$

Definicja 2.12. **Maksymalnym ryzykiem** estymatora $\hat{\varphi}_n$ w **tempie** ψ_n, na zbiorze Σ nazywamy liczbę:

$$r(\hat{\varphi}_n, \psi_n) = \sup_{\varphi \in \Sigma} R(\hat{\varphi}_n, \varphi, \psi_n).$$

Ponieważ wprowadziliśmy pojęcie tempa zbieżności to w naturalny sposób interesuje nas raczej asymptotyczne ryzyko minimaksowe (w tempie ψ_n):

$$\lim_{n \to \infty} \inf_{\hat{\varphi}_n} r(\hat{\varphi}_n, \psi_n).$$

Granica ta jednak, ze względu na występowanie tempa, wcale nie musi istnieć w danym modelu szczegółowym. Zastąpimy ją więc pojęciami *limes inferior* i *limes superior*, które zawsze istnieją i zażądamy od optymalnego tempa spełnienia słabszego warunku ograniczoności ciągów *liminf* oraz *limsup* ryzyk minimaksowych[17].

[16] Nazwa własna Autora. W źródłowej pracy [12] pomija się słowa „w tempie ψ_n". Wydaje się jednak właściwie zaznaczenie w nazwie w jakim kontekście funkcjonuje tutaj podejście minimaksowe.

[17] Zob. też uwagi w pracy [8], na s. 4-5.

Definicja 2.13. Ciąg dodatni ψ_n jest nazywany **niższym tempem zbieżności** (z ang. *lower rate of convergence*) dla zbioru Σ w pseudometryce $d(\cdot,\cdot)$, gdy:

$$\exists\, C_o > 0: \lim_{n\to\infty}\inf \inf_{\hat{\varphi}_n} r(\hat{\varphi}_n, \psi_n) \geq C_o.$$

Własność ta, jeśli zachodzi, ma charakter negatywny. Oznacza bowiem, że estymatory parametru φ nie zbiegają do niego zbyt szybko (w sensie wyżej opisanego asymptotycznego ryzyka minimaksowego). Zwróćmy też uwagę, że niższe tempo zbieżności nie musi być zdefiniowane jednoznacznie, może istnieć wiele ciągów ψ_n spełniających powyższy warunek (zob. [12], s. 48).

Definicja 2.14. Ciąg dodatni ψ_n^* jest nazywany **minimaksowym tempem zbieżności** dla zbioru Σ w pseudometryce $d(\cdot,\cdot)$, gdy jest niższym tempem zbieżności oraz:

$$\exists\, \hat{\varphi}_n^*\ \exists\, C_1 > 0: \lim_{n\to\infty}\sup r(\hat{\varphi}_n^*, \psi_n^*) \leq C_1.$$

Estymator spełniający tę własność również jest nazywamy **optymalnym w sensie asymptotycznego ryzyka minimaksowego** (por. def. 2.5). Niestety, **minimaksowe tempo zbieżności** również nie musi być wyznaczone jednoznacznie, lecz z dokładnością do stałej[18].

Dla wyjaśnienia znaczenia owej optymalności rozważmy ciąg nierówności:

$$0 < C_o \leq \lim_{n\to\infty}\inf \inf_{\hat{\varphi}_n} r(\hat{\varphi}_n, \psi_n^*) \leq \lim_{n\to\infty}\sup \inf_{\hat{\varphi}_n} r(\hat{\varphi}_n, \psi_n^*) \leq \lim_{n\to\infty}\sup r(\hat{\varphi}_n^*, \psi_n^*) \leq C_1.$$

Są one prawdziwe wprost z definicji odpowiednich wielkości. Przy założeniu istnienia asymptotycznego ryzyka estymatora $\hat{\varphi}_n^*$ oraz ryzyka minimaksowego otrzymujemy:

$$0 < C_o \leq \lim_{n\to\infty} \inf_{\hat{\varphi}_n} r(\hat{\varphi}_n, \psi_n^*) \leq \lim_{n\to\infty} r(\hat{\varphi}_n^*, \psi_n^*) \leq C_1.$$

Widać więc, że granice zostały tu zastąpione *limes inferior i superior*, a odpowiednie ciągi są przynajmniej ograniczone, skoro nie wiadomo czy powyższe granice istnieją. W szczególności asymptotyczne ryzyko estymatora $\hat{\varphi}_n^*$ jest ograniczone. Zauważmy też, że granice te (jeśli istnieją) nie są równe zero dla optymalnego tempa zbieżności. Można jednak wykazać (przy odpowiednich założeniach odnośnie do funkcji straty), że dla tempa „wolniejszego" granica ta wynosi zero, natomiast dla tempa „szybszego" zmierza do nieskończoności. W tym właśnie sensie tempo minimaksowe jest optymalne. W ramach modelu szczegółowego wyjaśnimy co oznacza pojęcie tempa „wolniejszego" lub „szybszego".

[18] Oznacza to, że istnieją funkcje straty takie, że ciąg $\psi_n = L\psi_n^*$, gdzie $L > 0$ również spełnia to kryterium optymalności (dla potęgowej i wskaźnikowej funkcji straty zob. [12], s. 49).

3. SZCZEGÓŁOWY MODEL STATYSTYCZNY

Przypomnijmy (zob. wstęp), że przedmiotem naszej analizy jest zbiór możliwości produkcyjnych[19] postaci:

$$T = \{(x, y) \in R_{+0}^{p+q} : \text{z } x \text{ można wyprodukować } y\}.$$

Element tego zbioru będziemy nazywać wykonalnym planem produkcyjnym i utożsamiać z jednostką produkcyjną lub obiektem gospodarczym, który go zrealizował. Pomocniczo zdefiniujemy jeszcze **granicę produkcyjną**[20] zbioru T:

$$\partial T = \{(x, y) \in T : \forall \theta \in (0,1) : (\theta x, y) \notin T, \forall \underline{\theta} > 1 : (x, \underline{\theta} y) \notin T\}.$$

Zbiór T jest **nieznany** i w związku z tym może być potraktowany jako parametr podlegający estymacji pod warunkiem zdefiniowania odpowiedniego modelu statystycznego.

Składa się on z szeregu założeń i jest szczególnym przypadkiem ogólnego schematu modelowania omówionego wcześniej.

Założenie 1’ Zbiór[21] $\mathcal{P}(R_{+0}^{p+q})$, z określoną pseudometryką miary Lebesgue’a symetrycznej różnicy zbiorów:

$$\forall A, B \in \mathcal{P}(R_{+0}^{p+q}) : \lambda(A \Delta B) = \lambda((A \backslash B) \cup (B \backslash A)),$$

jest **przestrzenią parametrów**.

Pseudometryka ta będzie mierzyć jak blisko nieznanego zbioru T (parametru) jest wartość jego estymatora. Oczywiście istnieją inne pseudometryki, które można by w tym celu wykorzystać (zob. zakończenie oraz praca [10], s. 478). Wybór tej właśnie jest podyktowany rezultatami teoretycznymi, które będą omawiane w niniejszej pracy.

Kolejne założenie będzie służyć wyodrębnieniu podzbioru $\mathcal{P}(R_{+0}^{p+q})$ wyrażającego naszą informację o nieznanym parametrze T.

Założenie 2’ T jest podrodziną $\mathcal{P}(R_{+0}^{p+q})$, której elementy są:

a. zbiorami zwartymi;

b. spełniają warunek tzw. swobodnej dyslokacji nakładów i produktów[22] tzn.:

$$\forall \underline{x} \geq x, \underline{y} \leq y [(x, y) \in T \Rightarrow (\underline{x}, \underline{y}) \in T];$$

---

[19] R_{+0}^{p+q} to iloczyn kartezjański p + q przedziałów postaci $[0,\infty)$.

[20] Za pracą [21], s. 421. Zob. też przypis 3. Zbiór ten opisywany jest też jako część wspólna zbioru T i domknięcia jego dopełnienia (w sensie topologicznym nie należy więc mylić go z pojęciem brzegu zbioru).

[21] Jest to rodzina wszystkich podzbiorów danego zbioru zwana **zbiorem potęgowym**.

[22] Nierówności między wektorami rozumiane „po współrzędnych”.

c. w przypadku braku nakładów nie można niczego wyprodukować[23]:

$$x = \mathbf{0},\ y \neq \mathbf{0} \Rightarrow (x, y) \notin T.$$

W razie potrzeby będziemy też zakładać, że:
d. T jest wypukły tzn.:

$$\forall x, y \in T \forall \alpha \in [0,1]: \alpha x + (1-\alpha) y \in T.$$

O tym, w jaki sposób generowane są obserwacje mówi następne założenie modelowe (jest to połączenie zał. 3 i 4 z ogólniejszego schematu):

Założenie 3' Próba statystyczna $\mathcal{X}_n = \left((x_i, y_i) \in R_{+0}^{p+q}, i = 1, \ldots, n\right) \in \left(\left(R_{+0}^{p+q}\right)^n; \mathcal{B}\left(R_{+0}^{p+q}\right)^n\right)$ jest realizacją ciągu niezależnych wektorów losowych (X_i, Y_i) o tym samym rozkładzie, każdy o gęstości (względem miary Lebesgue'a): $f_T(x, y): R_{+0}^{p+q} \to R$, ciągłej na nośniku[24] T.

UWAGA: Z niezależności wektorów losowych wynika, że gęstość łączna wyraża się wzorem:

$$f_T^n(\mathcal{X}_n) = \prod_{i=1}^{n} f_T(x_i, y_i).$$

Dla większej czytelności, w literaturze przedmiotu pomija się indeks n przy oznaczeniach gęstości, rozkładu prawdopodobieństwa oraz miary Lebesgue'a, która pełni tu rolę miary ν_n z zał. 4 ogólniejszego schematu modelowania.

Założenie 3' implikuje w szczególności, że:

$$\forall (x, y) \in \partial T : f_T(x, y) > 0 \text{ i ciągła w tym punkcie w kierunku wnętrza[25] T,}$$

$$P_T((x_i, y_i) \in T) = 1, \text{ dla dowolnego } i = 1, \ldots, n.$$

Jest to warunek wystarczający zgodności estymatorów rozważanych w dalszej części pracy. Oznaczający niezerową szansę zaobserwowania obiektu w dowolnym otoczeniu zbioru ∂T. Z punktu widzenia teorii mikroekonomicznej rynków konkurencyjnych jest to rozsądne założenie, gdyż oznacza, że w długim okresie firmy nieefektywne będą opuszczać rynek (za pracą [21], s. 427). Zakładamy ponadto, iż realizacje wektorów jako obserwowalne są planami wykonalnymi.

---

[23] Zapis $y \neq 0$ oznacza, iż przynajmniej jedna współrzędna wektora y jest różna od zera.

[24] Nośnik gęstości to zbiór argumentów, na którym przyjmuje ona wartości dodatnie.

[25] Wnętrze zbioru T ozn. intT to największy (w sensie inkluzji) zbiór otwarty zawarty w T. Pojęcie ciągłości w kierunku wnętrza zaś oznacza spełnienie warunku:

$$\forall \varepsilon > 0 \forall \{x_k, y_k : k \in N\} \subset \operatorname{int} T : (x_k, y_k) \to (x, y) \exists k_\varepsilon \forall k \geq k_\varepsilon \left| f(x_k, y_k) - f(x, y) \right| < \varepsilon.$$

Mamy określoną rodzinę gęstości $\{f_T(\cdot)\}_{T \in \mathcal{T}}$ na przestrzeni $\left(R_{+0}^{p+q}\right)^n$. Para $\left(\left(R_{+0}^{p+q}\right)^n, \{f_T(\cdot)\}_{T \in \mathcal{T}}\right)$ jest więc **modelem statystycznym** (przestrzenią statystyczną). Warto też podkreślić, że jest to model **nieparametryczny**, ponieważ nie istnieje skończenie wymiarowa parametryzacja $\mathcal{T}$.

W celu przeprowadzenia użytecznego wnioskowania statystycznego w oparciu o ten model ponownie określimy co rozumiemy pod pojęciem estymatora zbioru T (por. def. 2.1 ogólniejszego schematu).

Definicja 3.1. **Estymatorem** zbioru T nazywamy dowolne odwzorowanie mierzalne $\hat{T}:\left(R_{+0}^{p+q}\right)^n \to \mathcal{P}\left(R_{+0}^{p+q}\right)$.

Zwróćmy uwagę, że wartość estymatora nie musi być zbiorem z rodziny $\mathcal{T}$, do której należy nieznany parametr. W powyższych przestrzeniach mierzalnych przyjmujemy σ-algebry zbiorów borelowskich.

Uszczegółowimy teraz pojęcie funkcji ryzyka (por. def. 2.11) będące podstawą omówionego już kryterium minimaksowego optymalności estymatorów i tempa zbieżności.

Funkcją ryzyka $R(\cdot,\cdot,\cdot)$ w tempie $n^{-\alpha}$ nazywamy odwzorowanie

$$R(\hat{T}, T, n^{-\alpha}) = E_T\left[n^{\alpha} \cdot \lambda(\hat{T} \Delta T)\right],$$

gdzie[26]:
E_T – wartość oczekiwana rozkładu o gęstości $f_T(x, y)$ (zakładamy, że istnieje),
$n^{-\alpha}$, dla $\alpha > 0$ – przyjmujemy szczególną postać ciągów opisujących tempo zbieżności. Wraz ze wzrostem α, będziemy mówić, iż tempo zbieżności rośnie (jest **szybsze**).

Odpowiednie definicje, związane z funkcją ryzyka, zostały już wprowadzone przy okazji omawiania modelu ogólniejszego. Przystąpimy więc do przedstawienia dwóch wybranych propozycji[27] estymatorów zbioru możliwości produkcyjnych T na podstawie próby $\mathcal{X}_n$.

Definicja 3.2. **Estymatorem FDH zbioru T** jest:

$$T_{FDH}(\mathcal{X}_n) = \left\{(x, y) \in R_{+0}^{p+q}: \exists i \in \{1, \ldots, n\}: x \geq x_i, y \leq y_i\right\}.$$

Przypomnijmy, że wartość tego estymatora nie musi należeć do $\mathcal{T}$. Jednak dalsze rezultaty, dotyczące własności tego estymatora, uzyskano przy założeniu, że $T_{FDH}(\mathcal{X}_n) \in \mathcal{T}$. Zbiór ten jest najmniejszym zbiorem zawierającym obserwacje i spełniającym zał. 2'b o swobodnej dyslokacji (z ang. *free disposal*, stąd jego nazwa). Został wprowadzony do literatury przedmiotu w roku 1984 (praca [5]). Jest to zbiór domknięty, ograniczony ze względu na produkty, jednak formalnie nieograniczony ze względu

26 Za funkcję straty przyjęto wartość bezwzględną, stąd wobec nieujemności pseudometryki, nie występuje ona *explicite* w definicji funkcji ryzyka.

27 Istnieją inne propozycje (zob. zakończenie) nie są one jednak związane bezpośrednio z metodami DEA i FDH.

na nakłady. Nie spełnia więc w pełni zał. 2'a. Można jednak założyć, iż poszczególne nakłady nie mogą rosnąć nieograniczenie, czyli przyjąć, że:

$$\exists M > 0 : T_{FDH}(\mathcal{X}_n) \subset \left\{(x, y) \in [0, M]^{p+q} : \exists i \in \{1, \ldots, n\} : x \geq x_i, y \leq y_i\right\}.$$

Wtedy $T_{FDH}(\mathcal{X}_n)$, jako podzbiór domknięty zbioru zwartego[28], spełnia zał. 2'a.

Twierdzenie 3.1. Zbiór $T_{FDH}(\mathcal{X}_n)$ spełnia zał. 2'c.

Dowód: Załóżmy nie wprost, że zał. 2'c nie jest spełnione, wtedy:

$$\exists i \in \{1, \ldots, n\} : \mathbf{0} \geq x_i, y \leq y_i \text{ oraz } y \neq 0.$$

Wobec kierunków nierówności i nieujemności wielkości nakładów oznaczałoby to, iż:

$$\exists i \in \{1, \ldots, n\} : x_i = \mathbf{0}, y_i \neq 0 (*).$$

Z drugiej strony, zał. 3'implikuje:

$$P((x_i, y_i) \in T) = 1,$$

gdzie zbiór T spełnia zał. 2'c. Stoi to w sprzeczności z (*). c.b.d.o.

Łatwo również wykazać, że w większości praktycznych sytuacji zbiór ten nie jest wypukły, czyli nie spełnia zał. 2'd (zob. ilustracja). Przypomnijmy jednak, iż zał. 2'd został wprowadzony opcjonalnie. Oznacza to, że w sytuacji gdy podejrzewamy, iż zbiór możliwości produkcyjnych może nie być wypukły (np. niepodzielne jednostki nakładów i produktów, czy występowanie globalnie rosnącego efektu skali) warto zastosować ten estymator.

Przyjmując, że zbiór możliwości produkcyjnych spełnia zał. 2'd (wypukłość, czyli wykonalność „pośrednich" planów produkcyjnych) istnieje potrzeba użycia innego estymatora.

Definicja 3.3. **Estymatorem DEA[29] zbioru T** jest:

$$T_{DEA}(\mathcal{X}_n) = \left\{(x, y) \in R_{+0}^{p+q} : \exists \lambda_i \geq 0 : y \leq \sum_{i=1}^{n} \lambda_i \cdot y_i, x \geq \sum_{i=1}^{n} \lambda_i \cdot x_i, \sum_{i=1}^{n} \lambda_i = 1\right\}.$$

[28] Twierdzenie to można znaleźć np. w pracy [9], s. 164.

[29] Tę wersję estymatora DEA wprowadzono, w kontekście miar efektywności Farrella, w 1984 roku w pracy [2]. Istnieją inne wersje, które pojawiły się zarówno wcześniej, jak i później (zob. np. [15], s. 57-58).

Przy założeniu wypukłości T, estymator ten ma nieco lepsze własności niż $T_{FDH}(\mathcal{X}_n)$. Jednak użycie estymatora $T_{DEA}(\mathcal{X}_n)$, gdy zał. 2'd nie jest spełnione, prowadzi do braku jego zgodności (zob. tw. 4.1d). Ponownie przyjmuje się, że:

$$\exists M > 0 : T_{DEA}(\mathcal{X}_n) \subset \left\{(x, y) \in [0, M]^{p+q} : \exists \lambda_i \geq 0 : y \leq \sum_{i=1}^{n} \lambda_i \cdot y_i,\ x \geq \sum_{i=1}^{n} \lambda_i \cdot x_i,\ \sum_{i=1}^{n} \lambda_i = 1\right\}.$$

4. ILUSTRACJA I WŁASNOŚCI WPROWADZONYCH ESTYMATORÓW

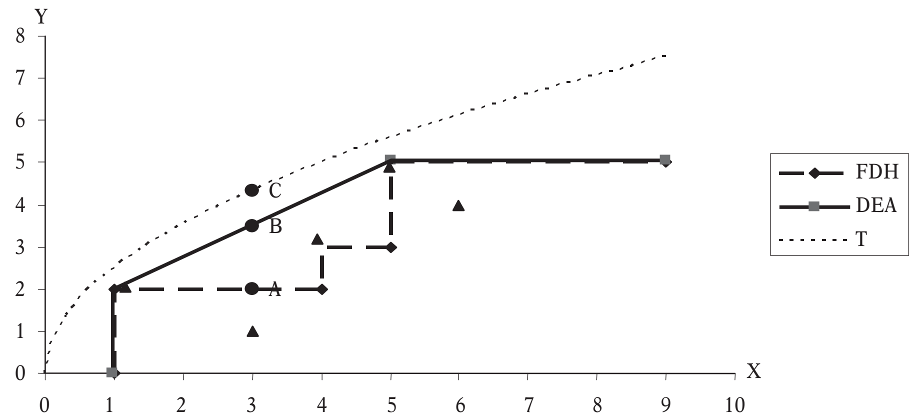
Załóżmy przykładowo, iż mamy w próbie pięć obiektów, które z jednego nakładu wytwarzają jeden produkt[30]. Odpowiednie dane, dotyczące wartości nakładu i produktu dla poszczególnych jednostek, zebrano w tabeli 1.

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.



Rysunek 1. Ilustracja dotycząca konstrukcji rozważanych estymatorów

Źródło: opracowanie własne.

[30] Umożliwi to ich ilustrację na płaszczyźnie w kartezjańskim układzie współrzędnych. Niewielka liczność próby (mimo asymptotycznych własności rozważanych estymatorów) nie zmniejsza ogólności przedstawionej ilustracji graficznej.

Zgodnie z zapisem, wprowadzonym na początku części trzeciej pracy, obiekty są przedstawiane za pomocą wektora wartości nakładu i produktu. Próbę można więc zapisać w następujący sposób: $\mathcal{X}_5 = \{(1,2); (3,1); (4,3); (5,5); (6,4)\}$. Wszystkie jednostki produkcyjne należą do nieznanego zbioru możliwości produkcyjnych T, który podlega estymacji (obszar ograniczony od góry drobną linią przerywaną na rysunku 1). Obszary, na rysunku 1, ograniczone od góry szarą oraz czarną linią łamaną to odpowiednio zbiory $T_{DEA}(\mathcal{X}_5)$ i $T_{FDH}(\mathcal{X}_5)$. Są to realizacje estymatorów zbioru T uzyskane dla naszej próby.

Zauważmy, że spełnione są inkluzje: $T_{FDH}(\mathcal{X}_5) \subseteq T_{DEA}(\mathcal{X}_5) \subseteq T$. Jest to ogólna prawidłowość wynikająca wprost z odpowiednich definicji, do której powrócimy przy okazji omawiania własności estymatorów. Zwróćmy ponadto uwagę na prostotę wyznaczania realizacji estymatorów zbioru możliwości produkcyjnych na podstawie próby.

W pracy [11] (s. 7) dowiedziono twierdzenie, które potwierdza optymalność wprowadzonych estymatorów. Wnioskiem z tego twierdzenia są własności estymatorów zebrane poniżej.

Twierdzenie 4.1. Dla ($p \geq 1$ oraz $q = 1$) lub ($q \geq 1$ oraz $p = 1$)

a. ($p \geq 1$, $q = 1$) $\lambda(T_{FDH}(\mathcal{X}_n)\ \Delta T) = O_P(n^{-1/p+1})$,

$$\lambda(T_{DEA}(\mathcal{X}_n)\ \Delta T) = O_P(n^{-2/p+2}), \text{ gdy T jest zbiorem wypukłym.}$$

b. ($q \geq 1$, $p = 1$) $\lambda(T_{FDH}(\mathcal{X}_n)\ \Delta T) = O_P(n^{-1/q+1})$,

$$\lambda(T_{DEA}(\mathcal{X}_n)\ \Delta T) = O_P(n^{-2/q+2}), \text{ gdy T jest zbiorem wypukłym.}$$

c. $T_{FDH}(\mathcal{X}_n)$, $T_{DEA}(\mathcal{X}_n)$ są estymatorami zgodnymi, gdy T jest zbiorem wypukłym.

d. $T_{FDH}(\mathcal{X}_n)$ jest estymatorem zgodnym i $T_{DEA}(\mathcal{X}_n)$ jest estymatorem niezgodnym, gdy T nie jest zbiorem wypukłym.

e. $T_{DEA}(\mathcal{X}_n)$ ma szybsze[31] tempo zbieżności niż $T_{FDH}(\mathcal{X}_n)$, gdy T jest zbiorem wypukłym.

f. Im większa jest liczba odpowiednio produktów lub nakładów, tym wolniejsza jest zbieżność obu estymatorów (tzw. przekleństwo wymiarowości, z ang. *curse of dimensionality*).

g. Jeśli zbiór T jest wypukły, to $T_{DEA}(X_n)$ jest estymatorem optymalnym w sensie (asymptotycznego) ryzyka minimaksowego. Jego tempo zbieżności z punktów a i b jest wtedy minimaksowym tempem zbieżności.

h. Jeśli zbiór T nie jest wypukły, to $T_{FDH}(X_n)$ jest estymatorem optymalnym w sensie (asymptotycznego) ryzyka minimaksowego. Jego tempo zbieżności z punktów a i b jest wtedy minimaksowym tempem zbieżności.

[31] Przykładowo $2/(p+2) > 1/(p+1)$ (zob. tw. 4.1a). Wtedy tempo $n^{-2/(p+2)}$ nazywa się szybszym od $n^{-1/(p+1)}$.

Z tw. 4.1c oraz tw. 2.1 wynika ponadto, że rozważane estymatory zbiorów są asymptotycznie nieobciążone, przy założeniu wypukłości zbioru T. Skoro zachodzą inkluzje $T_{FDH}(\mathcal{X}_n) \subseteq T_{DEA}(\mathcal{X}_n) \subseteq T$ estymatory te nazywa się, w skończonej próbie, **obciążonymi do wewnątrz** (z ang. *inward biased*). Widać też, że estymator DEA jest „bliższy" w sensie inkluzji nieznanemu zbiorowi T, stąd jego szybsze tempo zbieżności (zob. tw. 4.1e). Zwróćmy uwagę, że nie uzyskano żadnych rezultatów dla przypadku ogólnego, gdy $p, q \geq 1$. W dowodzie wspomnianego twierdzenia z pracy [11] korzystano z faktu, że dla $p = 1$ lub $q = 1$ granicę zbioru możliwości produkcyjnych można wyrazić jako funkcję o odpowiednich własnościach (wynikających z własności zbioru *T* ujętych w zał. 2').

5. ZAKOŃCZENIE

Niniejsze opracowanie ma charakter teoretyczny i służy jako podstawa badań o charakterze symulacyjnym i empirycznym. Ogólny schemat modelu, przedstawiony w części drugiej, ma wiele różnych szczegółowych wariantów, służących różnym zastosowaniom (przegląd można znaleźć w pracy [12]). Przedmiotem dalszych badań autora będzie m.in. próba estymacji gęstości łącznej rozkładu wprowadzonego w założeniu 3'.

Szczegółowy model statystyczny oraz dwie propozycje estymatorów zbioru możliwości produkcyjnych, przedstawione w pracy, ukierunkowane są na cel główny badań autora. Jest nim, jak wspomniano we wstępie, rozszerzenie deterministycznych metod FDH i DEA w kierunku stochastycznym. Przy zastosowaniu podejścia probabilistycznego nie tracimy bowiem informacji uzyskanych za pomocą wersji deterministycznej, a zyskujemy możliwość opisu niepewności związanej z estymacją[32]. Nawet w tym ograniczonym zakresie istnieją jeszcze inne rezultaty i propozycje modelowe. Przykładowo, w pracy [10] wprowadzono w przestrzeni parametrów pseudometrykę zwaną dystansem Hausdorffa oraz nałożono dodatkowe warunki na funkcję straty. Rozważano jedynie estymator FDH (nie zakładano warunku wypukłości). Uzyskano podobne własności, lecz inną postać optymalnego tempa zbieżności[33]. Owa postać tempa jest charakterystyczna dla rozważanej pseudometryki również pod kątem wcześniejszych rezultatów uzyskanych dla zbiorów wypukłych (zob. [7], [19]). Ponadto, w ogólnym przypadku (przy braku założenia wypukłości), proponuje się również inną postać estymatora zbioru zwaną estymatorem Devroye-Wise'a[34] (zob. [6]). Szeroki przegląd literatury dotyczącej estymacji zbiorów można znaleźć np. w pracy [4].

Autor wykorzystał przedstawiony powyżej schemat wnioskowania statystycznego w swoich dwóch wcześniejszych pracach (uszczegółowiony pod kątem estymacji miar efektywności technicznej Farrella). Ważne byłoby również wykorzystanie go w celu estymacji gęstości łącznej rozkładu wprowadzonej w zał. 3'. Problem ten wiąże się

---

[32] Dotyczy to nie tylko zbioru możliwości produkcyjnych, lecz przede wszystkim estymacji miary efektywności technicznej Farrella, która stanowi główne zastosowanie obu metod (uzyskujemy miary rozproszenia i przedziały ufności dla wartości wspomnianej miary).

[33] Tempo postaci $\psi_n^* = [(\log n)/n]^{1/(s+1)}$, gdzie $s = p$ (dla $q = 1$) lub $s = q$ (dla $p = 1$).

[34] Jest to odpowiednik tzw. estymatorów jądrowych służących do estymacji gęstości prawdopodobieństwa.

ściśle z estymacją nieznanych stałych w asymptotycznych rozkładach próbkowych estymatorów miar efektywności technicznej (zob. prace [14], [16]).

Uniwersytet Ekonomiczny w Krakowie

LITERATURA

[1] Banker R.D., [1993], *Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation*, Management Science, Vol. 39 (No. 10), s. 1265-1273.
[2] Banker R.D., Charnes A., Cooper W.W., [1984], *Some models for estimating technical and scale inefficiencies in DEA*, Management Science, Vol. 30 (No. 9), s. 1078-1091.
[3] Chevalier J., [1976], *Estimation du support et du contenu du support d'une loi de probabilité*, Annals Inst. H. Poincaré Sec. B Vol. 12, s. 339-364.
[4] Cuevas A., Rodriguez-Casal A., [2003], *Set estimation: an overview and some recent developments*, in „Recent Advances and Trends in Nonparametric Statistics" (red. Akritas, Politis), s. 251-264 North-Holland, Amsterdam.
[5] Deprins D., Simar L., Tulkens H., [1984], *Measuring labor efficiency in post offices*, in „The Performance of Public Enterprises: Concepts and Measurements" (red. Marchand, Pestieau, Tulkens), s. 243-267, Amsterdam, North-Holland.
[6] Devroye L., Wise L.G., [1980], *Detection of abnormal behavior via nonparametric estimation of the support*, SIAM J.Appl. Math. Vol. 38, s. 480-488.
[7] Dümbgen L., Walther G., [1996], *Rates of convergence for random approximations of convex sets*, Adv. Appl. Prob. Vol. 28, s. 384-393.
[8] Ibragimov I.A., Khasminskii R.Z., [1981], *Statistical Estimation: Asymptotic Theory*, Springer-Verlag, New York.
[9] Kleiner W., [1986], *Analiza matematyczna*, tom I, PWN Warszawa.
[10] Korostelev A.P., Simar L., Tsybakov A.B., [1995], *Efficient estimation of monotone boundaries*, The Annals of Statistics Vol. 23, s. 476-489.
[11] Korostelev A.P., Simar L., Tsybakov A.B., [1995], *On estimation of monotone and convex boundaries*, Publications de l'Institut de Statistique des Universités de Paris XXXIX Vol. 1, s. 3-18.
[12] Korostelev A.P., Tsybakov A.B., [1993], *Minimax Theory of Image Reconstruction*, Springer-Verlag, New York.
[13] Ombach J., [1993], *Wstęp do rachunku prawdopodobieństwa*, Wyd. UJ, Kraków.
[14] Pajor A., Prędki A., [2009], *Nieparametryczna estymacja miernika efektywności technicznej w ramach metody DEA*, Przegląd Statystyczny Vol. 56 (No. 3-4), s. 5-25.
[15] Prędki A., [2006], *Definiowanie globalnego i lokalnego efektu skali w ramach badania efektywności metodą DEA*, Przegląd Statystyczny Vol. 53 (No. 3), s. 57-72.
[16] Prędki A., *Propozycja opisu niepewności w ramach metod DEA i FDH*, wysłane do druku w Wyd. Uniwersytetu Wrocławskiego.
[17] Rényi A., Sulanke R., [1963], *Über die konvexe Hülle von n zufällig gewählten Punkten*, Z. Wahrschei nlichkeitstheorie Verw. Gebiete Vol. 2, s. 75-84.
[18] Rényi A., Sulanke R., [1964], *Über die konvexe Hülle von n zufällig gewählten Punktem II*, Z. Wahrsc heinlichkeitstheorie Verw. Gebiete Vol. 3, s. 138-147.
[19] Schneider R., [1988], *Random approximation of convex sets*, J. Microscopy Vol. 151, s. 211-227.
[20] Serfling R.J., [1991], *Twierdzenia graniczne statystyki matematycznej*, PWN Warszawa.
[21] Simar L., Wilson W., [2008], *Statistical Inference in Nonparametric Frontier Models: Recent Developments and Perspectives*, in "The Measurement of Productive Efficiency and Productivity Growth" (red. Fried, Lovell, Schmidt), Oxford University Press.

Praca wpłynęła do redakcji w czerwcu 2010 r.

ESTYMACJA ZBIORU MOŻLIWOŚCI PRODUKCYJNYCH W RAMACH FORMALNEGO MODELU STATYSTYCZNEGO

Streszczenie

W artykule przedstawiono ogólny model statystyczny oraz jedną, z jego szczegółowych wersji, służącą jako podstawa formalna estymacji zbioru możliwości produkcyjnych w ramach metod DEA i FDH. Przedstawiono własności estymatorów FDH i DEA oraz zilustrowano ich realizację w skończonej próbie. Wprowadzono elementy podejścia minimaksowego oraz wykorzystano pojęcie tempa zbieżności dla określenia definicji optymalności estymatora.

Słowa kluczowe: nieparametryczny model statystyczny, podejście minimaksowe, tempo zbieżności, zbiór możliwości produkcyjnych, estymacja zbiorów, metoda FDH, metoda DEA.

ESTIMATION OF PRODUCTION SET BASED ON FORMAL STATISTICAL MODEL

Summary

In the paper some general statistical model is presented and one of its particular version is exploited to formal estimate of the production set within the DEA and FDH methods. Properties of the FDH and DEA estimators are presented and their realizations for a finite sample are illustrated. Elements of the minimax approach are introduced and the rate of convergence is exploited to express the definition of asymptotic optimality of the estimators.

Key words: nonparametric statistical model, minimax approach, rate of convergence, production set, set estimation, FDH method, DEA method.