

POLSKA AKADEMIA NAUK
KOMITET STATYSTYKI I EKONOMETRII

PRZEGLĄD STATYSTYCZNY

STATISTICAL REVIEW

TOM 57

4

2010



DOM WYDAWNICZY ELIPSA
WARSZAWA 2010

WYDAWCA

Komitet Statystyki i Ekonometrii Polskiej Akademii Nauk

RADA REDAKCYJNA

Andrzej S. Barczak, Czesław Domański, Krzysztof Jajuga, Witold Jurek,
Michał Kolupa (Przewodniczący), Józef Pociecha, Danuta Strahl

KOMITET REDAKCYJNY

Mirosław Szreder (Redaktor Naczelny),
Magdalena Osińska (Zastępca Redaktora Naczelnego),
Marek Walesiak (Zastępca Redaktora Naczelnego),
Krzysztof Najman (Sekretarz Naukowy)

Wydanie publikacji dofinansowane przez Ministra Nauki i Szkolnictwa Wyższego

© Copyright by Komitet Statystyki i Ekonometrii PAN

Strona WWW „Przeglądu Statystycznego”:
<http://www.przegladstatystyczny.pan.pl>

Nakład 300 egz.



Realizacja wydawnicza:
Dom Wydawniczy ELIPSA,
ul. Inflancka 15/198, 00-189 Warszawa
tel./fax (22) 635 03 01, 635 17 85, e-mail: elipsa@elipsa.pl
www.elipsa.pl

ARTUR PRĘDKI

ESTYMACJA ZBIORU MOŻLIWOŚCI PRODUKCYJNYCH W RAMACH FORMALNEGO MODELU STATYSTYCZNEGO¹

1. WSTĘP

Niniejsze opracowanie stanowi kontynuację badań opisanych w pracach [14], [16], których celem było przedstawienie możliwych rozszerzeń metod FDH i DEA² 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 produkcyjnej³. 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 minimaxowego. 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. Zaczniemy 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 pseudometryki⁴.

Kolejne założenie definiuje pojęcie wektora obserwacji.

Założenie 2. Wektor obserwacji (próba statystyczna) 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 $\{P_\varphi^n, \varphi \in \Sigma \subset \Phi\}$ 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 $\xi_n: (X, \sigma_X, P) \rightarrow (X_n, \sigma_n)$ zwany **próbą losową**, którego rozkład jest przedmiotem zainteresowania⁵. Jego realizacja to właśnie wektor obserwacji 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:

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

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 $(X_n, \sigma_n) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ 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 $\hat{\varphi}_n: (X_n, \sigma_n) \rightarrow (\Phi, d(\cdot, \cdot))$ 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 borelowskich⁶ $\mathcal{B}(\Phi)$.

Przejdziemy teraz do problemu zdefiniowania tzw. funkcji ryzyka (z ang. *risk function*) będącej podstawą, sformułowanego w dalszym ciągu pracy, kryterium optymalności estymatorów zwanego minimaxowym. Na gruncie statystyki parametrycznej funkcja ryzyka jest zwykle postaci⁷

$$R(\hat{\varphi}_n, \varphi) = E_{\varphi}^n [w(\hat{\varphi}_n, \varphi)] = \int_{X_n} w(\hat{\varphi}_n, \varphi) dP_{\varphi}^n,$$

gdzie:

$w(\cdot, \cdot)$ – dowolna nieujemna funkcja zwana **funkcją straty**⁸ (z ang. *loss function*);

E_{φ}^n – wartość oczekiwana względem rozkładu P_{φ}^n (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:

$$R(\hat{\varphi}_n, \varphi) = E_{\varphi}^n [w(d(\hat{\varphi}_n, \varphi))].$$

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 nieznannej 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 istnieć⁹ (tzw. estymator **optymalny w sensie ryzyka minimaksowego**). Stąd zwykle ogranicza się klasę rozważanych estymatorów¹⁰ 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 $\lim_{n \rightarrow \infty} \inf_{\hat{\varphi}_n} r(\hat{\varphi}_n)$ nazywamy **asymptotycznym ryzykiem minimaksowym**.

Jeśli istnieje ta granica¹¹ i estymator $\hat{\varphi}_n^*$ takie, że:

$$\lim_{n \rightarrow \infty} \inf_{\hat{\varphi}_n} r(\hat{\varphi}_n) = \lim_{n \rightarrow \infty} r(\hat{\varphi}_n^*),$$

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:

$$\lim_{n \rightarrow \infty} E_{\varphi}^n [d(\hat{\varphi}_n, \varphi)] = 0.$$

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

$$\forall \varepsilon > 0: \lim_{n \rightarrow \infty} P_{\varphi}^n (d(\hat{\varphi}_n, \varphi) < \varepsilon) = \lim_{n \rightarrow \infty} P_{\xi_n} \{ \mathcal{X}_n \in X_n : d(\hat{\varphi}_n(\mathcal{X}_n), \varphi) < \varepsilon \} = 1.$$

Stosuje się oznaczenia: $\hat{\varphi}_n \xrightarrow{P} \varphi$ lub $d(\hat{\varphi}_n, \varphi) = o_P(1)$. Zwróćmy uwagę, że pseudometryka $d(\cdot, \cdot)$ pełni tu ponownie rolę odległości euklidesowej w „zwykłej” definicji zgodności dla modeli parametrycznych¹².

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**, gdy¹³:

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

Stosuje się oznaczenie¹⁴: $d(\hat{\varphi}_n, \varphi) = O_P(1)$.

Warunek zgodności jest tu najsilniejszy, o czym mówi poniższe twierdzenie¹⁵.

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 minimaxowego).

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:

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

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:

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

Twierdzenie 2.3. Jeśli $d(\hat{\varphi}_n, \varphi) = O_P(\psi_n)$ i $\forall n \in N: \underline{\psi}_n \geq \psi_n, \underline{\psi}_n \rightarrow 0$, 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 $R(\cdot, \cdot, \cdot)$ w tempie¹⁶ ψ_n nazywamy odwzorowanie określone następująco:

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

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):

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

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 minimaksowych¹⁷.

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:

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

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 minimaxowego). 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:

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

Estymator spełniający tę własność również jest nazywany **optymalnym w sensie asymptotycznego ryzyka minimaxowego** (por. def. 2.5). Niestety, **minimaksowe tempo zbieżności** również nie musi być wyznaczone jednoznacznie, lecz z dokładnością do stałej¹⁸.

Dla wyjaśnienia znaczenia owej optymalności rozważmy ciąg nierówności:

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

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

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

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 produkcyjnych¹⁹ postaci:

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

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' Zbiór²¹ $\mathcal{P}(R_{+0}^{p+q})$, z określoną pseudometryką miary Lebesgue'a symetrycznej różnicy zbiorów:

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

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:

- zbiorami zwartymi;
- spełniając warunek tzw. swobodnej dyslokacji nakładów i produktów²² 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 wyprodukować²³:

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

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 $\mathcal{X}_n = ((x_i, y_i) \in R_{+0}^{p+q}, i = 1, \dots, n) \in ((R_{+0}^{p+q})^n; \mathcal{B}(R_{+0}^{p+q})^n)$ jest realizacją ciągu niezależnych wektorów losowych (X_i, Y_i) o tym samym rozkładzie, każdy o gęstości (względem miary Lebesgue'a): $f_T(x, y): R_{+0}^{p+q} \rightarrow R$, ciągłej na nośniku²⁴ 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 v_n z zał. 4 ogólniejszego schematu modelowania.

Założenie 3' implikuje w szczególności, że:

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

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

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. $\text{int} T$ to największy (w sensie inkluzji) zbiór otwarty zawarty w T . Pojęcie ciągłości w kierunku wnętrza zaś oznacza spełnienie warunku:

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

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

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 $\hat{T}: (R_{+0}^{p+q})^n \rightarrow \mathcal{P}(R_{+0}^{p+q})$.

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

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

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

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

gdzie²⁶:

E_T – wartość oczekiwana rozkładu o gęstości $\mathbf{f}_T(x, y)$ (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 propozycji²⁷ estymatorów zbioru możliwości produkcyjnych T na podstawie próby $\mathcal{X}_n$.

Definicja 3.2. Estymatorem FDH zbioru T jest:

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

Przypomnijmy, że wartość tego estymatora nie musi należeć do $\mathbf{T}$. Jednak dalsze rezultaty, dotyczące własności tego estymatora, uzyskano przy założeniu, że $T_{\text{FDH}}(\mathcal{X}_n) \in \mathbf{T}$. 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 *explicitie* 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:

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

Wtedy $T_{FDH}(X_n)$, jako podzbiór domknięty zbioru zwartego²⁸, spełnia zał. 2'a.

Twierdzenie 3.1. Zbiór $T_{FDH}(X_n)$ spełnia zał. 2'c.

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

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

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

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

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 DEA²⁹ zbioru T jest:

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

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 $T_{FDH}(X_n)$. Jednak użycie estymatora $T_{DEA}(X_n)$, gdy zał. 2'd nie jest spełnione, prowadzi do braku jego zgodności (zob. tw. 4.1d). Ponownie przyjmuje się, że:

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

4. ILUSTRACJA I WŁASNOŚCI WPROWADZONYCH ESTYMATORÓW

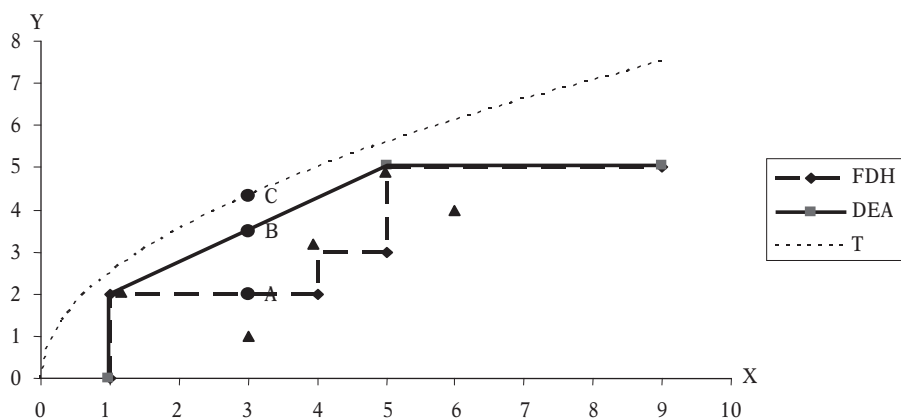
Założmy przykładowo, iż mamy w próbie pięć obiektów, które z jednego nakładu wytwarzają jeden produkt³⁰. Odpowiednie dane, dotyczące wartości nakładu i produktu dla poszczególnych jednostek, zebrano w tabeli 1.

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: $X_5 = \{(1,2); (3,1); (4,3); (5,5); (6,4)\}$. 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 $T_{DEA}(X_5)$ i $T_{FDH}(X_5)$. Są to realizacje estymatorów zbioru T uzyskane dla naszej próby.

Zauważmy, że spełnione są inkluzje: $T_{FDH}(X_5) \subseteq T_{DEA}(X_5) \subseteq T$. 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. ($p \geq 1, q = 1$) $\lambda(T_{FDH}(X_n) \Delta T) = O_p(n^{-1/p+1})$,

$\lambda(T_{DEA}(X_n) \Delta T) = O_p(n^{-2/p+2})$, gdy T jest zbiorem wypukłym.

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

$\lambda(T_{DEA}(X_n) \Delta T) = O_p(n^{-2/q+2})$, gdy T jest zbiorem wypukłym.

c. $T_{FDH}(X_n)$, $T_{DEA}(X_n)$ są estymatorami zgodnymi, gdy T jest zbiorem wypukłym.

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

e. $T_{DEA}(X_n)$ ma szybsze³¹ tempo zbieżności niż $T_{FDH}(X_n)$, 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 minimaxowego. Jego tempo zbieżności z punktów a i b jest wtedy minimaxowym tempem zbieżności.

h. Jeśli zbiór T nie jest wypukły, to $T_{FDH}(X_n)$ jest estymatorem optymalnym w sensie (asymptotycznego) ryzyka minimaxowego. Jego tempo zbieżności z punktów a i b jest wtedy minimaxowym 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 $T_{FDH}(X_n) \subseteq T_{DEA}(X_n) \subseteq T$ 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 estymacją³². 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 zbieżności³³. 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’a³⁴ (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 nieznanymi 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., Prędkie A., [2009], *Nieparametryczna estymacja miernika efektywności technicznej w ramach metody DEA*, Przegląd Statystyczny Vol. 56 (No. 3-4), s. 5-25.
- [15] Prędkie A., [2006], *Definiowanie globalnego i lokalnego efektu skali w ramach badania efektywności metodą DEA*, Przegląd Statystyczny Vol. 53 (No. 3), s. 57-72.
- [16] Prędkie 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. Wahrscheinlichkeitstheorie Verw. Gebiete Vol. 2, s. 75-84.
- [18] Rényi A., Sulanke R., [1964], *Über die konvexe Hülle von n zufällig gewählten Punkten II*, Z. Wahrscheinlichkeitstheorie 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.

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 minimaxowego oraz wykorzystano pojęcie tempa zbieżności dla określenia definicji optymalności estymatora.

Słowa kluczowe: nieparametryczny model statystyczny, podejście minimaxowe, 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.

CZESŁAW DOMAŃSKI

UWAGI O TESTACH JARQUE'A-BERA

1. MIARY SKOŚNOŚCI I KURTOZY

W literaturze statystycznej prezentuje się wiele miar skośności i spłaszczenia (kurtozy). Występują różne podziały tych miar, często jako kryterium uwzględnia się wielkość wpływu na ich wartość pojedynczych obserwacji nietypowych. Miary te często są stosowane do budowy statystyk testowych weryfikujących hipotezę o skośności rozkładów (por. Domański [4] i [5]).

Przypuśćmy, że mamy n niezależnych obserwacji $x_1, x_2, \dots, x_n$ zmiennej losowej X .

Dla n -elementowej próby określamy moment centralny k -tego rzędu:

$$m_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k, k = 2, 3, \dots \quad (1)$$

$$\text{Statystyki } \sqrt{b_1} = \frac{m_3}{\sqrt{m_2^3}} \text{ i } b_2 = \frac{m_4}{m_2^2} \quad (2)$$

są nieobciążonymi estymatorami parametrów

$$\sqrt{\beta_1} = \frac{\mu_3}{\sigma^3} \text{ i } \beta_2 = \frac{\mu_4}{\sigma^4}. \quad (3)$$

Jeśli zmienna losowa X ma rozkład normalny o parametrach μ i σ , to wiadomo, że $\mu_3 = 0$, $\mu_4 = 3\sigma^4$, stąd $\sqrt{\beta_1} = 0$ i $\beta_2 = 3$.

W artykule zostaną przedstawione własności testów weryfikujących hipotezy o skośności rozkładów zaproponowanych przez Jarque'a-Berego [7] oraz ich modyfikacji zaproponowanych przez Godfreya i Orme'a [6].

2. TESTY JARQUE'A-BEREGO I MODYFIKACJE

Test Jarque'a-Berego [7] oparty jest na miarach skośności i spłaszczenia (kurtozy) rozkładu, czyli jest testem omnibus, ponieważ uwzględnia jednocześnie odstępstwa od

normalności wywołane zarówno przez skośności i spłaszczenia. Statystyka ta najczęściej przedstawiona jest wzorem:

$$JB = n \left[\frac{(\sqrt{b_1})^2}{6} + \frac{(b_2 - 3)^2}{24} \right]. \quad (4)$$

Statystyka (4) przy założeniu prawdziwości hipotezy zerowej H_0 , że rozkład badany jest rozkładem normalnym, ma asymptotyczny rozkład χ^2 o dwóch stopniach swobody. Wydzielając pierwszy składnik, uwzględniający jedynie skośność rozkładu, dostajemy postać:

$$JB_1 = n \frac{(\sqrt{b_1})^2}{6}. \quad (5)$$

Statystyka ta przy założeniu prawdziwości hipotezy $H_0: \sqrt{\beta_1} = 0$ (współczynnik skośności równa się zero) ma asymptotyczny rozkład χ^2 o jednym stopniu swobody.

Wyodrębniając drugi składnik ze wzoru (4) otrzymujemy:

$$JB_2 = n \frac{(b_2 - 3)^2}{24}. \quad (6)$$

Statystyka (6) przy założeniu prawdziwości hipotezy zerowej $H_0: \beta_2 = 3$ (współczynnik spłaszczenia równa się 3) ma asymptotyczny rozkład χ^2 o jednym stopniu swobody.

W wielu pracach wykazano, że test oparty na statystyce (5) nie rozróżnia w sposób wiarygodny rozkładów symetrycznych i asymetrycznych w przypadku występowania wysokiej kurtozy. Istotną przyczyną występowania nieodpowiedniej własności tego testu jest fakt, że wzór na wariancję statystyki $\sqrt{b_1}$ został wyprowadzony przy założeniu wystąpienia niskiej kurtozy. Bera i John [2] wyraźnie stwierdzili, że statystyki (5) i (6) nie są poprawnymi testami skośności i kurtozy, ponieważ asymptotyczny rozkład tych statystyk jest wyprowadzony przy założonej normalności.

W szczególności wzór określający asymptotyczną wariancję $D^2(\sqrt{b_1}) = 6/n$ wykorzystany we wzorze (5) będzie niepoprawny. Na przykład, przy dużej kurtozie będziemy mieli do czynienia z proporcjonalnie większą liczbą obserwacji nietypowych, co często występuje w praktyce przy badaniu rozkładów stóp zwrotu instrumentów finansowych (por. Piontek [9]). Pearson [8] wykazał, że wychodząc od rozkładów z krótkimi ogonami i przesuwając się w kierunku rozkładów o długich ogonach, momenty będą rosły coraz szybciej ze względu na ekstremalne wartości pochodzące z ogonów rozkładów prawdopodobieństwa. W konsekwencji wariancja statystyki (5) będzie rosła i dla rozkładów o grubych ogonach formuła $D^2(\sqrt{b_1}) = 6/n$ będzie powodowała niedoszacowanie wariancji. W przypadku dużej kurtozy, będziemy odrzucać hipotezę zerową zakładając symetrię zbyt często.

Godfrey i Orme [6] zaproponowali odpowiednią modyfikację poszerzając zakres zastosowań tego testu zwłaszcza do badania rozkładów o grubych ogonach (por. także Bera, Premarante 2001).

Znalezienie formuły dla wariancji statystyk $\sqrt{b_1}$ oraz b_2 (por. (2)), które będą poprawne w przypadku błędnej specyfikacji modelu, umożliwia wykorzystanie rodziny rozkładów Pearsona oraz zastosowania modyfikacji White'a (1987).

Zmodyfikowana statystyka przyjmuje postać:

$$RS = n \frac{(\sqrt{b_1})^2}{9 + m_6 m_2^{-3} - 6 m_4 m_2^{-2}} \quad (7)$$

gdzie m_i oznaczają odpowiednie i -te momenty zwykłe rozkładu zmiennej losowej X .

Statystyka (7) przy założeniu prawdziwości hipotezy H_0 ma również asymptotyczny rozkład χ^2 o jednym stopniu swobody.

3. PORÓWNANIE TESTÓW OPARTYCH NA STATYSTYCE $(\sqrt{b_1})$

W celu zbadania własności rozważanych testów opartych na statystykach (5) i (7) wygenerowaliśmy metodą Monte Carlo dane z rozkładu normalnego (10000 powtórzeń). Uzyskaliśmy wynik świadczący o tym, że rozmiar zmodyfikowanego testu opartego na statystyce RS jest na ogół bliższy założonym poziomom istotności w szczególności dla $\alpha = 1\%$ (por. tabl. 1).

W tabeli 1 przedstawione są wyniki dotyczące rozmiaru testu dla rozkładu Studenta (t_7) o 7 stopniach swobody. W tym przypadku standardowy test niedoszacowuje wariancję, prowadząc do błędnego zbyt częstego odrzucania hipotezy $H_0: \sqrt{\beta} = 0$.

Wartość wyrażenia $(9 + \mu_6 \eta_2^{-2} - 6\mu_4 \eta_2^{-2})/n = 104/n$ we wzorze (7) jest zdecydowanie większa niż standardowa wielkość. Prawdopodobieństwo odrzucenia hipotezy H_0 dla testu zmodyfikowanego jest również w tym przypadku bliskie nominalnym poziomom 5% i 1%. Proces generujący dane z rozkładu beta (2,2), ma charakter platykuryczny $\beta_2 < 3$ oraz symetryczny. W tym przypadku standardowy test cechuje się znacznie bardzo małymi oszacowaniami wartościami rozmiaru. Zmodyfikowany test tym razem również ma rozmiar bliższy nominalnym poziomom 5% i 1%.

Dla rozkładu Laplace'a, który cechuje symetrię oraz leptokurozę ($\beta_2 > 0$) standardowy test odrzuca prawdziwą hipotezę zbyt często, podczas gdy zmodyfikowany test utrzymuje swój rozmiar.

Tabela 2 przedstawia rezultaty symulacji dotyczące mocy rozważanych testów. Dla dwóch pierwszych rozkładów, które generują dane z dodatnią asymetrią (χ_4^2 oraz beta 1,2) zarówno test standardowy, jak i zmodyfikowany cechuje się bardzo wysoką mocą, przy czym dla rozkładu beta (1,2) wyższą moc posiada test zmodyfikowany.

Analiza wyników prezentowanych w tablicy 1 i 2 potwierdza tezę, że standardowy test oparty na statystyce określonej wzorem (5) daje błędne wyniki w przypadku występowania dużej kurtozy. W konsekwencji powoduje to zbyt częste, błędne odrzu-

cenie prawdziwej hipotezy zerowej mówiącej o symetrii. Test zmodyfikowany oparty na statystyce (7) daje znacznie lepsze rezultaty. Ponadto cechuje się bardzo dobrymi właściwościami odnoszącymi się do jego rozmiaru i mocy.

Tabela 1

Oszacowania rozmiarów testów skośności (asymetrii) na podstawie 10000 powtórzeń

Wielkość próby	Test			
	standardowy (JB ₁)		zmodyfikowany (RS)	
	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$
Rozkład $N(0,1)$				
30	3,29	0,83	3,88	0,59
50	4,35	1,07	4,43	0,71
100	4,88	1,12	4,55	0,81
200	4,83	0,89	4,51	0,63
250	4,74	0,91	4,68	0,81
300	4,96	0,89	4,96	0,73
400	5,17	1,06	4,81	0,90
500	4,99	0,91	5,00	0,85
600	4,88	1,05	4,73	0,80
700	4,62	0,94	4,42	0,65
800	4,92	1,05	5,00	0,82
900	4,52	0,89	4,43	0,78
1000	5,18	0,87	4,98	0,88
Rozkład Studenta t_7				
30	15,36	7,71	3,08	0,39
50	20,93	11,99	2,66	0,35
100	27,85	17,45	3,02	0,27
200	33,48	22,27	3,44	0,31
250	34,72	23,29	2,81	0,20
300	36,79	24,98	3,28	0,39
400	37,72	26,16	3,45	0,34
500	39,91	27,27	3,50	0,36
600	39,65	28,12	3,38	0,37
700	41,44	29,09	3,67	0,43
800	41,72	29,94	3,50	0,30
900	43,05	30,91	3,65	0,33
1000	43,29	31,47	3,68	0,57

cd. tabeli 1

Wielkość próby	Test			
	standardowy (JB ₁)		zmodyfikowany (RS)	
	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$
Rozkład Beta (2,2)				
30	0,17	0,01	3,05	0,63
50	0,11	0,00	3,03	0,53
100	0,17	0,00	3,39	0,42
200	0,11	0,00	2,79	0,41
250	0,04	0,00	2,75	0,44
300	0,07	0,00	3,17	0,37
400	0,10	0,00	2,96	0,47
500	0,14	0,00	3,16	0,53
600	0,08	0,00	3,04	0,50
700	0,11	0,00	2,73	0,39
800	0,08	0,00	2,80	0,42
900	0,13	0,00	2,89	0,41
1000	0,06	0,00	2,98	0,45
Rozkład Laplace'a ($\lambda = 2$)				
30	26,18	15,49	2,85	0,19
50	33,47	21,58	2,28	0,14
100	39,74	27,50	2,74	0,18
200	44,07	31,77	3,05	0,32
250	46,07	33,10	3,06	0,16
300	47,49	35,26	3,41	0,29
400	47,73	35,08	3,41	0,34
500	48,56	36,13	3,59	0,31
600	49,30	36,77	3,65	0,32
700	49,95	37,05	3,68	0,33
800	49,34	37,49	4,00	0,46
900	50,83	38,22	3,96	0,44
1000	50,65	38,76	4,07	0,52

Źródło: obliczenia własne.

Tabela 2

Oszacowania rozmiarów testów skośności metodą Monte Carlo oparte na 10000 powtórzeń

Wielkość próby	Test			
	standardowy (JB_1)		zmodyfikowany (RS)	
	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$
Rozkład χ_4^2				
30	60,90	39,24	85,02	58,70
50	87,79	70,81	97,51	91,58
100	99,68	97,91	99,99	99,90
200	100	100	100	100
250	100	100	100	100
300	100	100	100	100
400	100	100	100	100
500	100	100	100	100
600	100	100	100	100
700	100	100	100	100
800	100	100	100	100
900	100	100	100	100
1000	100	100	100	100
Rozkład Beta (1,2)				
30	11,97	2,36	49,93	21,04
50	27,33	6,70	73,39	49,87
100	66,90	31,92	96,21	88,18
200	97,98	84,17	99,96	96,97
250	99,57	95,09	100	100
300	99,95	98,43	100	100
400	100	99,93	100	100
500	100	100	100	100
600	100	100	100	100
700	100	100	100	100
800	100	100	100	100
900	100	100	100	100
1000	100	100	100	100

cd. tabeli 2

Wielkość próby	Test			
	standardowy (JB ₁)		zmodyfikowany (RS)	
	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$
Rozkład Beta (2,1)				
30	12,78	36,87	2,15	14,85
50	26,96	61,62	7,03	35,59
100	68,04	92,60	31,55	76,87
200	97,78	99,87	85,19	99,01
250	99,68	100	94,82	99,86
300	99,93	100	98,64	99,99
400	100	100	99,96	100
500	100	100	100	100
600	100	100	100	100
700	100	100	100	100
800	100	100	100	100
900	100	100	100	100
1000	100	100	100	100

Źródło: obliczenia własne.

4. UWAGI KOŃCOWE

Warto prześledzić własności testu opartego na statystyce określonej wzorem (6) oraz jego modyfikacji postaci:

$$RS^* = n \frac{(b_2 - 3)^2}{24 + 64b_1 + 8b_3}$$

gdzie: b_1 i b_2 dane są wzorem (2), $b_3 = \frac{m_3 m_5}{m_2^4}$,

przy czym m_i ($i = 1, 2, \dots, 5$) oznaczają momenty zwykłe zmiennej losowej.

Uniwersytet Łódzki

LITERATURA

- [1] Asai M., Dashzeveg U., [2006], *Distribution – Free Test for Symmetry with an Application to S&P Index Returns*, <http://home.soka.ac.jp/faculty/economics?DP/004.pdf>
- [2] Bera A.K., John S., [1983], *Test for multivariate normality with Pearson alternatives*, *Communication Statistics – Theory and Methods*, 12, s. 103-117.
- [3] Bera A., Premarante G., [2001], *Adjusting the Tests for Skewness and Kurtosis for Distributional Misspecifications*, http://www.business.uiuc.edu/Working_Papers/papers/01-0116.pdf
- [4] Domański Cz., [1990], *Testy statystyczne*, Państwowe Wydawnictwo Ekonomiczne, Warszawa.
- [5] Domański Cz., [2009], *Testy normalności oparte na momentach*, *Prace i Materiały Wydziału Zarządzania Uniwersytetu Gdańskiego* 4/2, s. 569-577.
- [6] Godfrey L., Orme C., [1991], *Testing for Skewness of Regression Distributions*, *Economic Letters* 37, s. 31-34.
- [7] Jarque C.M., Bera A.K., [1987], *A Tests of Observations and Regression Residuals*, *International Statistical Review*, vo. 55, s. 163-172.
- [8] Pearson E.S., [1963], *Some problems arising in approximating to probability distributions using moments*, *Biometrika* 50, s. 95-112.
- [9] Piontek K., [2007], *Pomiar i testowanie skośności rozkładów stóp zwrotu instrumentów finansowych*, *Prace Naukowe Akademii Ekonomicznej we Wrocławiu, Taksonomia* 14, s. 122-130.
- [10] White H., [1982], *Maximum likelihood estimation of misspecified models*, *Econometrica*, 50, s. 1-25.

Praca wpłynęła do redakcji w lipcu 2010 r.

UWAGI O TESTACH JARQUE'A-BERA

Streszczenie

W opracowaniu tym analizie symulacyjnej poddane zostały rozmiary testów skośności i kurtozy Jarque'a-Bera oraz ich modyfikacji. Eksperymenty symulacyjne dotyczą wybranych czterech rodzin rozkładów prawdopodobieństwa i obejmują 10000 powtórzeń każdy.

Słowa kluczowe: statystyczne testy skośności i kurtozy, rozmiar testu, eksperymenty symulacyjne

SOME REMARKS ON JARQUE-BERA TESTS

Summary

The paper involves a description of simulation experiments aimed at determining the size of Jarque-Bera tests for skewness and kurtosis, as well as selected modifications of those tests. All experiments refer to four different families of probability distributions and consist of one thousand runs each.

Key words: statistical tests for skewness and kurtosis, the size of a test, simulation experiments.

PIOTR WDOWIŃSKI*

POLITYKA GOSPODARCZA W MODELU CEN I KURSU WALUTOWEGO

1. WPROWADZENIE

W ostatnim dwudziestoleciu nastąpił w Europie i na świecie szybki wzrost poziomu integracji gospodarczej, szczególnie w odniesieniu do rynków finansowych. W państwach Unii Europejskiej (UE) ten proces stanowił fundament dla powstania Unii Gospodarczej i Walutowej (UGW). Powstanie UGW oznaczało zastąpienie walut narodowych wspólną walutą euro. Kompetencje narodowych banków centralnych w zakresie polityki pieniężnej przejął Europejski Bank Centralny (ECB). Powstanie UGW przyczyniło się m.in. do zwiększenia skali obrotów w handlu zagranicznym Polski, co ma wpływ na szybszy wzrost gospodarczy.

Członkostwo w UE określa obecną i przyszłą politykę kursową w Polsce. Wprowadzenie systemu kursów płynnych w Polsce było ważne z dwóch powodów. Po pierwsze, miało na celu umożliwienie prowadzenia przez Radę Polityki Pieniężnej polityki bezpośredniego celu inflacyjnego¹. Po drugie, miało umożliwić osiągnięcie przez złotego poziomu równowagi długookresowej, który byłby korzystny z punktu widzenia późniejszego przystąpienia do strefy euro.

Utrata autonomii w prowadzeniu polityki pieniężnej przez niezależny bank centralny na rzecz innej instytucji ponadnarodowej może mieć znaczne konsekwencje gospodarcze dla danego kraju. Są one tym poważniejsze, im większe różnice – gospodarcze i kulturowe – obserwuje się pomiędzy krajami, które zamierzają przystąpić do unii walutowej. W tym kontekście ważne jest określenie zakresu oddziaływania instrumentów polityki gospodarczej – fiskalnej i pieniężnej – na wzrost gospodarczy. Jeśli gospodarki nie mają solidnych podstaw w sferze realnej, to załamania koniunktury na świecie, w tym gwałtowne spadki popytu i fluktuacje rynku finansowego, szybko pobudzają deprecjację walut tych krajów².

Wybór właściwego reżimu walutowego stanowi jeden z głównych problemów polityki gospodarczej. W artykule zostanie omówione zagadnienie efektywności reżimu walutowego kursów zmiennych. Przedstawione rozważania można rozumieć w kontek-

* Uniwersytet Łódzki oraz Urząd Komisji Nadzoru Finansowego. Artykuł wyraża osobiste poglądy autora i nie jest oficjalnym stanowiskiem UKNF. Autor pragnie podziękować Recenzentowi za krytyczne uwagi, które przyczyniły się do poprawy treści artykułu.

¹ Różne aspekty polityki pieniężnej przez pryzmat reguły Taylora są przedstawione w pracy [2].

² Zagadnienia ryzyka walutowego i modele kryzysów walutowych zostały przedstawione w pracach [10], [11]. Por. również [17], [9].

ście przyszłego członkostwa Polski w UGW. Oczekiwane członkostwo Polski w systemie ERM2 oraz przystąpienie do UGW oznacza stopniowe usztywnianie kursu złotego aż do całkowitego wprowadzenia wspólnej waluty. Tym samym system kursów zmieniających się płynnie można traktować jako „system utraconych korzyści”.

W artykule dokonano próby teoretycznej analizy skutków funkcjonowania małej gospodarki otwartej w systemie kursu zmiennego³. Podstawą wnioskowania jest model gospodarki, zorientowany podażowo, funkcjonujący w warunkach kursu zmiennego, należący do nurtu ekonomii matematycznej. Model stanowi rozwinięcie standardowych modeli według Mundella-Fleminga-Dornbuscha. Jego konstrukcja pozwala na opisanie najważniejszych zależności ekonomicznych. Model można przedstawić w postaci dynamicznego układu równań różniczkowych pierwszego rzędu, którego rozwiązanie analityczne umożliwia przeprowadzenie analiz symulacyjnych. Celem budowy modelu jest ułatwienie zrozumienia konsekwencji prowadzenia polityki gospodarczej – pieniężnej, fiskalnej i płacowej – w systemie zmiennego kursu walutowego (szerzej na ten temat por. np. [20]).

Wśród prac stanowiących motywację dla przedstawionego modelu należy wyróżnić takie opracowania, jak: [7], [14], [3], [18], [13], [4], [1], [15], [21]. Przedmiotem zainteresowania w wymienionych opracowaniach jest zagadnienie stabilności makroekonomicznej małej gospodarki otwartej. Część z nich dotyczy UGW i UE w kontekście optymalnych reżimów walutowych oraz wewnętrznych i zewnętrznych konsekwencji utworzenia wspólnego obszaru walutowego.

2. MODEL CEN I KURSU WALUTOWEGO

W teorii parytetu siły nabywczej (PPP) przyjmuje się, że kurs walutowy zależy w największym stopniu od cen towarów podlegających wymianie handlowej (handlowych), nie zaś od cen towarów niepodlegających wymianie handlowej (niehandlowych) (por. np. [20]). Przedstawiona analiza dotyczyć będzie modelu zorientowanego podażowo, funkcjonującego w warunkach kursu zmiennego, w którym przyjęto podział na sektory towarów handlowych i niehandlowych.

W artykule bieżący kurs walutowy oznaczono symbolem s , który został zdefiniowany jako krajowa cena waluty obcej, np. cena euro w złotych. Konsekwencją przyjęcia takiej definicji kursu jest terminologia dotycząca wzrostu lub spadku tej ceny. Wzrost kursu oznacza deprecjację waluty krajowej względem waluty obcej, spadek zaś kursu odpowiednio aprecjację. Ponadto nominalny kurs rynkowy wyższy od umownie przyjętego kursu równowagi oznacza „podwartościowość” waluty krajowej i konsekwentnie „nadwartościowość” w przeciwnym przypadku.

Niech całkowity popyt⁴ y_T^d na towary handlowe (T) jest funkcją krajowego dochodu y , nominalnej stopy procentowej i , poziomu konkurencyjności gospodarki krajowej względem gospodarki światowej (realnego kursu walutowego) c , relacji cen towarów

³ Szerszą dyskusję na temat związków pomiędzy kursem walutowym, cenami i wzrostem gospodarczym można znaleźć np. w pracach [8], [19], [23], [12] i [22].

⁴ Małymi literami oznaczono logarytmy naturalne odpowiednich zmiennych.

handlowych p_T i niehandlowych p_N oraz zmiennej $u^d = \{g, y^*\}$, która stanowi ocenę zakłócenia popytowego. Wśród zakłóceń można wyróżnić zmiany polityki fiskalnej g lub zmiany popytu zagranicznego y^* . Niech krzywa IS ma postać:

$$y_T^d = \delta y - \sigma i + \alpha c - \gamma(p_T - p_N) + u^d, \quad (1)$$

gdzie: $\delta, \sigma, \alpha, \gamma$ – parametry modelu. Parametr δ odzwierciedla efekt mnożnikowy, uwzględniający konsumpcję prywatną, inwestycje i import. Należy zauważyć, że kluczowe dla modelu są wzajemne powiązania pomiędzy sektorami, $p_T - p_N$ (por. [13]).

Niech funkcja podaży towarów handlowych ma postać:

$$y_T = -\varphi(w_T - p_T), \quad (2)$$

gdzie: w_T – egzogeniczna płaca przeciętna w sektorze towarów handlowych, natomiast φ – parametr prędkości dostosowań.

Niech funkcja całkowitego popytu na krajowe towary niehandlowe (N) ma postać:

$$y_N^d = \delta y - \sigma i + \gamma(p_T - p_N) + u^d. \quad (3)$$

Podaż towarów niehandlowych można opisać wzorem:

$$y_N = -\varphi(w_N - p_N), \quad (4)$$

gdzie: w_N – egzogeniczna płaca przeciętna w sektorze towarów niehandlowych.

W równaniach (1)-(4) zakłada się, że wrażliwość popytu i podaży w obu sektorach na zmiany dochodu, stopy procentowej, relacji cen i płac realnych jest jednakowa, tzn. przyjmuje się identyczne parametry.

Przyjęto również, że:

$$w_T = w_N = \bar{w}. \quad (5)$$

Warunek (5) oznacza, że przepływ siły roboczej pomiędzy sektorami prowadzi do wyrównania się stawek płac. Gdyby przyjąć, że sektor towarów handlowych jest poddany warunkom konkurencji ze strony czynników egzogenicznych, wynikających z międzynarodowej swobody przepływu pracy, natomiast sektor towarów niehandlowych dostosowuje się do warunków sektora towarów handlowych, to warunek (5) przyjąłby postać:

$$w_N = \omega w_T, \quad 0 < \omega < 1. \quad (6)$$

Warunek (6) oznacza, że wzrost wydajności pracy w sektorze towarów handlowych przenosiłby się do sektora towarów niehandlowych, jednak nie w sposób proporcjonalny.

Równowagę na rynku pieniężnym opisano za pomocą krzywej LM:

$$m^s - p = \kappa y - \lambda i, \quad (7)$$

gdzie $m^s - p$ oznacza realną podaż pieniądza, przy czym deflatorem jest zagregowany indeks cen p . Po prawej stronie związku (7) występuje funkcja popytu na pieniądz, w tym popytu transakcyjnego, związanego z poziomem dochodu y i spekulacyjnego, związanego z nominalną stopą procentową i . Przyjęto, że podaż pieniądza składa się wyłącznie z krajowego kredytu d dla gospodarki, stąd $m^s = d$.

Niech zagregowany indeks cen p stanowi ważoną sumę indeksów cen towarów handlowych p_T i niehandlowych p_N :

$$p = \theta p_T + (1 - \theta) p_N, \quad 0 < \theta < 1, \quad (8)$$

gdzie: θ – udział towarów handlowych w strukturze koszyka konsumpcji.

Dynamikę indeksów cen dla towarów handlowych i niehandlowych można wyrazić za pomocą krzywych Phillipsa:

$$\dot{p}_T = \eta(y_T^d - y_T), \quad (9)$$

$$\dot{p}_N = \eta(y_N^d - y_N), \quad (10)$$

gdzie: η – prędkość dostosowania cen w warunkach nierównowagi na rynkach towarów handlowych i niehandlowych, przy czym symbol kropki nad zmienną oznacza pierwszą pochodną względem czasu.

Parametr η odgrywa szczególną rolę. Odzwierciedla wszystkie te czynniki, które łącznie wyznaczają poziom elastyczności cen w gospodarce, przy czym $\eta \rightarrow \infty$ oznacza warunki elastyczności doskonałej.

Ze względu na przyjęte w modelu ograniczenia w procesie dostosowania cen ($\eta \rightarrow 0$), zmienny kurs walutowy stanowi mechanizm dostosowawczy w sytuacji, gdy gospodarka jest poddawana zakłóceniom. Wówczas kurs zmienny odgrywa istotną rolę w przywracaniu równowagi w gospodarce. W długim okresie następuje pełne dostosowanie cen, dochód narodowy powraca do długookresowego stanu równowagi i pozostaje neutralny ze względu na impulsy pieniężne lub fiskalne.

Przyjęto, że oczekiwania względem kursu walutowego kształtują się regresywnie (por. [16]):

$$E(\dot{s}) = -\psi(s - \bar{s}), \quad (11)$$

gdzie:

s – nominalny kurs walutowy, $\bar{s}$ – długookresowy kurs walutowy, ψ – parametr dostosowań do stanu równowagi, E – operator oczekiwań. Formuła oczekiwań (11) może być zgodna z formułą przewidywania doskonałego, o ile poprawnie przyjmie się wartość parametru ψ (por. [6]).

Przyjęto również, że zachodzi „niezabezpieczony” parytet stóp procentowych:

$$E(\dot{s}) = i - i^*, \quad (12)$$

gdzie: i^* – zagraniczna nominalna stopa procentowa (egzogeniczna), która dla ułatwienia równa się zeru.

Niech poziom konkurencyjności c gospodarki krajowej względem gospodarki światowej przedstawia odchylenie nominalnego kursu walutowego od jego długookresowego stanu równowagi:

$$c = s - \bar{s}. \quad (13)$$

W równaniu (13) przyjmuje się, że gospodarka krajowa uzyskuje przewagę konkurencyjną nad gospodarką światową wówczas, gdy nominalny bieżący kurs walutowy przewyższa długookresowy kurs równowagi. Zgodnie z warunkami (11), (12) i (13), konkurencyjność można opisać zależnością:

$$c = -\frac{1}{\psi}i. \quad (14)$$

Przy warunku (14), równanie (1) można zapisać w postaci:

$$y_T^d = \delta y - \mu i - \gamma(p_T - p_N) + u^d, \quad (15)$$

gdzie: $\mu \equiv \sigma + \frac{\alpha}{\psi}$.

Biorąc pod uwagę założenie, że kurs długookresowy jest kształtowany przez ceny towarów handlowych, przyjmuje się dla niego postać warunku PPP:

$$\bar{s} = p_T - p_T^*, \quad (16)$$

gdzie: p_T^* – poziom cen zagranicznych (egzogenicznych), przyjęty dla wygody obliczeń jako równy zeru. Wówczas można wyrazić kurs walutowy jako funkcję dynamiki cen:

$$s - p_T = -\frac{1}{\psi}i. \quad (17)$$

Ostatecznie przyjęto również, że:

$$y = y_T + y_N. \quad (18)$$

Model można przedstawić w postaci układu dwóch równań różniczkowych pierwszego rzędu względem cen, tj. $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}$, gdzie $\mathbf{x}' = [p_T \ p_N]$. Układ ma postać:

$$\begin{bmatrix} \dot{p}_T \\ \dot{p}_N \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} p_T \\ p_N \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad (19)$$

gdzie:

$$\begin{aligned} a_{11} &= \eta\varphi(\delta - 1) - \eta\gamma - \frac{\eta\mu}{\lambda}(\varphi\kappa + \theta), \quad a_{12} = \eta(\delta\varphi + \gamma) - \frac{\eta\mu}{\lambda}(\varphi\kappa - \theta + 1), \\ a_{21} &= \eta(\varphi\delta + \gamma) - \frac{\eta\sigma}{\lambda}(\varphi\kappa + \theta), \quad a_{22} = \eta(\varphi\delta - \varphi - \gamma) - \frac{\eta\sigma}{\lambda}(\varphi\kappa - \theta + 1), \\ b_1 &= \left[\frac{2\eta\varphi\kappa\mu}{\lambda} + \eta\varphi(1 - 2\delta) \right] \bar{w} + \frac{\eta\mu}{\lambda}d + \eta u^d, \\ b_2 &= \left[\frac{2\eta\varphi\kappa\sigma}{\lambda} + \eta\varphi(1 - 2\delta) \right] \bar{w} + \frac{\eta\sigma}{\lambda}d + \eta u^d, \end{aligned}$$

gdzie: $\mu \equiv \sigma + \frac{\alpha}{\psi}$. Przyjmuje się, że układ (19) jest stabilny⁵. W stanie spoczynku układu przyjmuje się, że $p_T(0) = p_N(0) = p(0) = 0$. Macierz **A** posiada dwie ujemne wartości własne. W tym przypadku warunkami stabilności systemu (19) są: dodatni wyznacznik macierzy **A** i ujemny ślad tej macierzy.

Przyjęto następujące założenia dotyczące efektywności instrumentów polityki makroekonomicznej:

– deprecjacja waluty krajowej ma niewielki wpływ na wzrost eksportu i tym samym na wzrost popytu na krajową produkcję z uwagi na silną transmisję wzrostu kursu na wzrost poziomu cen krajowych, stąd parametr α jest niewielki, przy czym jego dodatnia wartość oznacza, że zachodzi warunek Marshalla-Lernera-Robinson;

– gospodarka jest silnie uzależniona od importu, stąd dochodowa elastyczność popytu na import jest wysoka i tym samym efekt mnożnikowy w postaci parametru δ jest niewielki;

– ceny są sztywne, co wyraża niska wartość parametru η ;

– całkowity popyt na produkcję jest wrażliwy na zmiany stopy procentowej i tym samym na zmiany inwestycji, stąd przyjmuje się wysoką wartość parametru σ .

W tab. 1 podano skalibrowane wartości parametrów przyjętych w symulacjach.

Tabela 1

Wartości parametrów oraz początkowych stanów zmiennych sterujących modelu

Parametr	δ	σ	α	γ	ϕ	θ	η	κ	γ	ψ	u^d	d	$\bar{w}$
Wartość	0,2	1	0,1	0,1	0,1	0,5	0,2	1	1	0,1	0	0	0

Źródło: opracowanie własne.

⁵ Na temat zagadnień stabilności systemowej układów dynamicznych por. [5].

W podanym środowisku makroekonomicznym model jest stabilny. Przeprowadzona analiza wrażliwości, polegająca na niewielkich zmianach parametrów, nie wykazała dużych zmian w dynamice systemu i nie wpłynęła na jego stabilność (por. również [21]). W następnym punkcie zostaną przeprowadzone analizy symulacyjne.

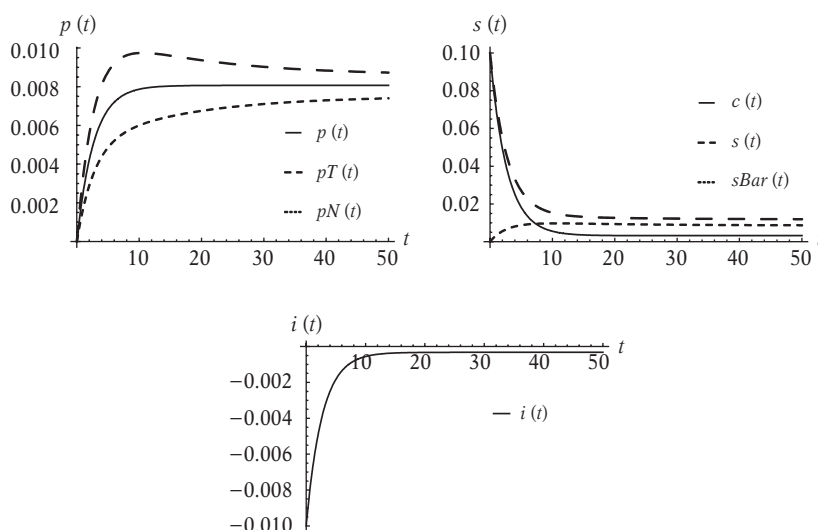
3. SYMULACJA SKUTKÓW POLITYKI GOSPODARCZEJ

W scenariuszach pokazano ścieżkę rozwoju, kolejno: krajowego indeksu cen (p, p_T, p_N), kursu walutowego ($c, s, \bar{s}$), nominalnej stopy procentowej (i). Na osi poziomej przedstawiono czas t wyrażony w kwartałach. Indeksy cen, kursy walutowe i stopa procentowa są wyrażone w logarytmach naturalnych. Dla tych zmiennych na osi pionowej wartości dodatnie (ujemne) odpowiadają stanom wyższym (niższym) o określony procent od stanu spoczynku układu.

Rozwiązanie analityczne modelu pozwala na przeprowadzenie szeregu analiz symulacyjnych (wykonanych w pakiecie *Mathematica*). W dalszej części zostaną przedstawione scenariusze symulacyjne dla ekspansywnej polityki pieniężnej i fiskalnej oraz scenariusz zmiany stawki płac.

3.1. POLITYKA PIENIĘŻNA

Zakłócenie pieniężne polegało na zaburzeniu zmiennej $d = 0,01$ (wzrost wolumenu kredytu o 1%). Skutki polityki pieniężnej przedstawia rys. 1.



Rysunek 1. Skutki ekspansywnej polityki pieniężnej

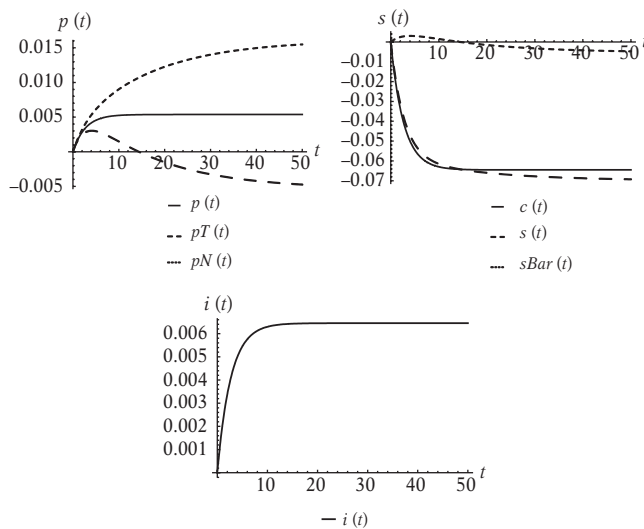
W modelu ekspansywna polityka pieniężna prowadzi do natychmiastowej deprecjacji waluty krajowej, przy czym kurs wykazuje cechy „przestrzelenia” na skutek spadku stopy procentowej i krótkookresowej sztywności cen. W dyskusji nad założeniami modelu pokazano bowiem, że poziom konkurencyjności c jest malejącą funkcją stopy procentowej. Efekt „przestrzelenia” rozpoznał i opisał R. Dornbusch. „Przestrzelenie” oznacza, że w krótkim okresie wzrost kursu walutowego przewyższa wzrost podaży pieniądza. Tę nadmierną deprecjację nazywa się „przestrzeleniem” kursu walutowego s ponad długookresowy kurs równowagi $\bar{s}$. Efekt „przestrzelenia” wynika z mniejszej elastyczności cen na tle elastyczności stopy procentowej i kursu walutowego. Ponadto można zauważyć, że w krótkim okresie realizuje się wzrost cen i spadek kursu walutowego. Oznacza to, że o ile w długim okresie zachodzi parytet PPP, to przejście do tego położenia równowagi odbywa się przy cenach i kursie zmieniających się w przeciwnych kierunkach. Stąd parytet PPP występuje w modelu jako zależność długookresowa z możliwymi odchyleniami kursu nominalnego od parytetu w krótkim okresie (por. [6]).

Jak widać, początkowy wzrost konkurencyjności c powoduje wzrost popytu na rynku towarów handlowych (T), przewyższający wzrost popytu na rynku towarów niehandlowych (N). W konsekwencji obserwuje się gwałtowniejszy wzrost poziomu cen p_T niż p_N . Biorąc pod uwagę, że parytet PPP jest spełniony ze względu na poziom cen p_T , natomiast kurs równowagi $\bar{s} = p_T$, to ich wyższy poziom przynosi osłabienie przewagi konkurencyjnej wobec warunków, w których kurs $\bar{s}$ miałby się dostosować do poziomu cen p_N lub p . W konsekwencji nie wystąpiły warunki podtrzymania popytu na krajową produkcję.

3.2. POLITYKA FISKALNA

Zakłócenie fiskalne polegało na zaburzeniu zmiennej $u^d = 0,01$. Skutki polityki fiskalnej przedstawia rys. 2.

Ekspansja fiskalna (dodatnie zakłócenie popytowe) prowadzi do wzrostu stopy procentowej i do aprecjacji waluty krajowej. Spadkowi poziomu konkurencyjności c towarzyszy spadek popytu na rynku towarów handlowych i kompensacyjny spadek poziomu cen p_T . W konsekwencji, wobec wzrostu poziomu cen p_N , wywołanego impulsem popytowym, spadek cen p_T ma wpływ na ograniczenie skali inflacji w gospodarce. Widać zatem wyraźnie, że wzrost stopy procentowej wpływa na ograniczenie inflacji, gdyż następuje kompensacyjny spadek poziomu cen na rynku towarów handlowych, związany z ograniczeniem popytu na tym rynku poprzez kanał stóp procentowych i kursu walutowego. Należy zauważyć, że im wyższy parametr θ , tym większe ograniczenie skali inflacji.

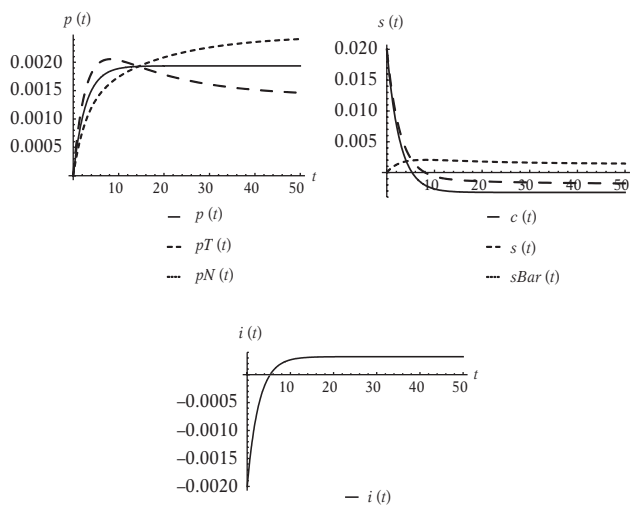


Rysunek 2. Skutki ekspansywnej polityki fiskalnej

Źródło: opracowanie własne.

3.3. POLITYKA PŁACOWA

Zakłócenie płacowe polegało na zaburzeniu stawki płac $\bar{w} = 0,01$ (por. rys. 3).



Rysunek 3. Wzrost nominalnej stawki płac

Źródło: opracowanie własne.

Wzrost egzogenicznej stawki płac prowadzi do spadku podaży towarów na obydwu rynkach (T , N) oraz w całej gospodarce. Skutkuje to spadkiem popytu na pieniądź i konsekwentnym spadkiem stopy procentowej w celu przywrócenia równowagi na rynku pieniężnym. W wyniku zmiany popytu na krajową produkcję, zostają uruchomione procesy inflacyjne. Spadkowi stopy procentowej towarzyszy początkowo nominalna i realna deprecjacja walutowa z widocznym efektem „przestrzelenia”. W okresie późniejszym, w konsekwencji stopniowego wygasania wzrostu zagregowanego poziomu cen następuje spadek nominalnego kursu walutowego s oraz wzrost kursu równowagi $\bar{s}$, związany ze wzrostem poziomu cen p_T . W tych warunkach kurs nominalny s ustalił się poniżej kursu długookresowego $\bar{s}$ i dało się zaobserwować spadek kursu realnego c .

Jeśli zatem utożsamiać wzrost egzogenicznej stawki płac $\bar{w}$ ze wzrostem wydajności pracy w obydwu sektorach, to można zauważyć, że realny kurs walutowy c jest malejącą funkcją wydajności pracy, co jest zgodne z założeniami modelu Balassy-Samuelsona.

4. ZAKOŃCZENIE

W artykule przedstawiono dwusektorowy model cen i kursu walutowego, należący do nurtu ekonomii matematycznej, funkcjonujący w warunkach reżimu kursu zmiennego. Model zapisano w postaci układu dwóch równań różniczkowych pierwszego rzędu względem cen towarów handlowych i niehandlowych. Za pomocą modelu pokazano główne kanały transmisji polityki gospodarczej – pieniężnej, fiskalnej i płacowej – na ceny i realny kurs walutowy.

W szczególności, po pierwsze, ekspansywna polityka pieniężna prowadziła do natychmiastowej deprecjacji waluty krajowej, kurs zaś wykazał cechy „przestrzelenia” na skutek początkowego spadku stopy procentowej i krótkookresowej sztywności cen. Tym samym nie wystąpiły warunki dla podtrzymania popytu na krajową produkcję, gdyż nastąpił szybszy wzrost cen towarów handlowych niż niehandlowych. Po drugie, ekspansja fiskalna prowadziła do wzrostu stopy procentowej i do aprecjacji waluty krajowej, co powodowało ograniczenie inflacji na skutek kompensacyjnego spadku poziomu cen handlowych poprzez kanał stóp procentowych i kursu walutowego. Po trzecie, wzrost wydajności pracy w długim okresie wiązał się z realną aprecjacją waluty krajowej.

Uniwersytet Łódzki

LITERATURA

- [1] Aarle B. van, [1996], *Essays on Monetary and Fiscal Policy Interactions: Applications to EMU and Eastern Europe*, Center for Economic Research, Tilburg University, Tilburg.
- [2] Baranowski P., [2008], *Reguła Taylora i jej rozszerzenia*, „Gospodarka Narodowa” 7-8.
- [3] Buiter W.H., Miller M., [1982], *Real Exchange Rate Overshooting and the Output Cost of Bringing Down Inflation*, „European Economic Review” 18(1), 85-123.

- [4] Cardia E., [1991], *The Dynamics of a Small Open Economy in Response to Monetary, Fiscal, and Productivity Shocks*, „Journal of Monetary Economics” 28, 411-434.
- [5] Chiang A., [1984], *Fundamental Methods of Mathematical Economics*, McGraw-Hill, New York.
- [6] Dornbusch R., [1976], *Expectations and Exchange Rate Dynamics*, „Journal of Political Economy” 84(6), 1161-1176.
- [7] Fleming J.M., [1962], *Domestic Financial Policies under Fixed and under Floating Exchange Rates*, „IMF Staff Papers” 12, 369-380.
- [8] Karadeloglou P., Chobanov G., Delakorda A., Milo W., Wdowiński P., [2001], *The Exchange Rate, Prices and the Supply Response under Transition: A Simulation Study*, [w:] Papazoglou C., Pentecost E.J. (eds.), *Exchange Rate Policies, Prices and Supply-Side Response. A Study of Transitional Economies*, Palgrave, New York.
- [9] Małecki W., Sławiński A., Piasecki R., Żuławska U., [2001], *Kryzysy walutowe*, PWN, Warszawa.
- [10] Milo W., Kozera Z., [2003], *Uwagi o pomiarze ryzyka walutowego*, [w:] *Preferencje a Ryzyko*, Wydawnictwo Akademii Ekonomicznej w Katowicach.
- [11] Milo W., Kozera Z., [2004], *O ryzyku kryzysu walutowego*, [w:] Milo W., Wdowiński P. (red.), *Prognozowanie rynków finansowych*, „Acta Universitatis Lodziensis. Folia Oeconomica” 177, Wydawnictwo Uniwersytetu Łódzkiego, Łódź, 57-78.
- [12] Milo W., Malaczewski M., Szafrński G., Ulrichs M., Wośko Z., [2010], *Stabilność rynków kapitałowych a wzrost gospodarczy*, PWN, Warszawa.
- [13] Montiel P.J., [1987], *Output and Unanticipated Money in the Dependent Economy Model*, „IMF Staff Papers” 34(2), 228-259.
- [14] Mundell R.A., [1963], *Capital Mobility and Stabilization Policy under Fixed and Flexible Exchange Rates*, „Canadian Journal of Economics and Political Science” 29, 475-485.
- [15] Papazoglou C., Pentecost E.J., [2001], *Output Dynamics in Transition Economies under Alternative Exchange Rate Regimes*, [w:] Papazoglou C., Pentecost E.J. (eds.), *Exchange Rate Policies, Prices and Supply-Side Response. A Study of Transitional Economies*, Palgrave, New York, 11-23.
- [16] Pentecost E.J. [1993], *Exchange Rate Dynamics: A Modern Analysis of Exchange Rate Theory and Evidence*, E. Elgar, Aldershot, Hants, England.
- [17] Sławiński A., [2000], *Kryzysy walutowe a kierunki reformy międzynarodowego systemu finansowego*, „Bank i Kredyt” 7-8, 90-102.
- [18] Turnovsky S.J., [1986], *Monetary and Fiscal Policy under Perfect Foresight: A Symmetric Two-country Analysis*, „Economica” 53, 139-157.
- [19] Wdowiński P., [2005], *Financial Markets and Economic Growth in Poland: Simulations with an Econometric Model*, [w:] Milo W., Wdowiński P. (eds.), *Issues in Modeling, Forecasting and Decision-Making in Financial Markets*, „Acta Universitatis Lodziensis. Folia Oeconomica” 192, Wydawnictwo Uniwersytetu Łódzkiego, Łódź, 27-53.
- [20] Wdowiński P., [2010], *Modele kursów walutowych, rozprawa habilitacyjna*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.
- [21] Wdowiński P., Aarle B. van, [2001], *Economic Performance in Poland under Fixed and Flexible Exchange Rate Regimes*, [w:] Papazoglou C., Pentecost E.J. (eds.), *Exchange Rate Policies, Prices and Supply-Side Response. A Study of Transitional Economies*, Palgrave, New York, 140-156.
- [22] Welfe W. (red.), [2010], *Makroekonometryczny model gospodarki opartej na wiedzy*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.
- [23] Wośko Z., [2005], *Zredukowany model rynku finansowego – symulacje*, [w:] Milo W., Wdowiński P. (red.), *Wybrane zagadnienia ilościowych analiz ekonomiczno-finansowych*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.

POLITYKA GOSPODARCZA W MODELU CEN I KURSU WALUTOWEGO

Streszczenie

W artykule przedstawiono dwusektorowy model cen i kursu walutowego, należący do nurtu ekonomii matematycznej, funkcjonujący w warunkach reżimu kursu zmiennego. Model zapisano w postaci układu dwóch równań różniczkowych pierwszego rzędu względem cen towarów handlowych i niehandlowych. Za pomocą modelu pokazano główne kanały transmisji polityki gospodarczej – pieniężnej, fiskalnej i płacowej – na ceny i realny kurs walutowy w warunkach podziału produkcji na towary handlowe i niehandlowe.

Słowa kluczowe: ceny, kurs walutowy, polityka pieniężna i fiskalna, polityka płacowa

ECONOMIC POLICY IN THE MODEL OF PRICES AND EXCHANGE RATE

Summary

This paper presents a two-sector model of prices and exchange rate, belonging to the mainstream of mathematical economics. The model operates under floating exchange rate regime. The model is given in the system form of two first order differential equations with respect to prices of tradables and non-tradables. Using the model we show the main transmission channels of economic policy – monetary, fiscal and wage – on prices and real exchange rate in terms of production divided into tradables and non-tradables goods.

Key words: prices, exchange rate, monetary and fiscal policy, wage policy

EWA WĘDROWSKA

WYKORZYSTANIE ENTROPII SHANNONA I JEJ UOGÓLNIEN DO BADANIA ROZKŁADU PRAWDOPODOBIENSTWA ZMIENNEJ LOSOWEJ DYSKRETNEJ

1. WPROWADZENIE

Pojęcie entropii, które wywodzi się z termodynamiki znalazło zastosowanie w wielu dziedzinach nauki, poczynszyszy od teorii informacji poprzez mechanikę statystyczną, teorię chaosu do sieci neuronowych, lingwistyki matematycznej czy też taksonomii. Możliwość wykorzystania entropii do określania stopnia nieokreśloności badanych stanów czy też rozkładów zmiennych losowych dało przyczynek do określania różnych postaci entropii. Do teorii informacji pojęcie entropii wprowadził Shannon (1948), po czym w drugiej połowie ubiegłego wieku pojawiło się wiele uogólnień probabilistycznej miary shannonowskiej entropii.

W artykule przedstawione zostało pojęcie entropii Shannona wraz z jej wybranymi uogólnieniami: entropią Rényiego (1965) oraz entropią Havrda-Charvát-Daróczy-Tsallisa (1970). Entropia Rényiego jest wynikiem uogólnienia średniej Kołmogorowa-Nagumo, entropia Tsallisa jest natomiast pewną funkcją entropii Rényiego.

Własności entropii rozkładu zmiennej losowej typu skokowego wskazują na liczne związki miar koncentracji z entropią Shannona ([2], [3], [7], [9], [10], [16], [17], [23], [24]). Dlatego też w artykule zaproponowano wykorzystanie entropii Rényiego, będącej uogólnieniem entropii Shannona, do oceny stopnia nierównomierności dyskretnego rozkładu prawdopodobieństwa.

Celem teoretycznym artykułu jest przedstawienie pojęć i własności entropii zmiennej losowej dyskretnej umożliwiające zastosowanie jej w badaniu własności dyskretnych rozkładów zmiennych losowych. Celem aplikacyjnym natomiast jest wskazanie na możliwość wykorzystania miar entropii do określania stopnia koncentracji dyskretnego rozkładu prawdopodobieństwa.

2. ENTROPIA SHANNONA

Niech X będzie zmienną losową dyskretną o rozkładzie prawdopodobieństwa $\{p(x_1), p(x_2), \dots, p(x_n)\}$ z prawdopodobieństwami $\{p(x_i)\}$ spełniającymi warunek

$0 \leq p(x_i) \leq 1$, $\sum_{i=1}^n p(x_i) = 1$. W teorii informacji zdefiniowanie entropii poprzedzone

zostało sformułowaniem następujących postulatów, jakie powinna spełniać poszukiwana funkcja entropii $H_S(X) = H_S(p(x_1), p(x_2), \dots, p(x_n))$ ([4],[12]):

P1. Funkcja $H_S(p(x_1), p(x_2), \dots, p(x_n))$ powinna być ciągła względem wszystkich argumentów $p(x_i)$ ($i = 1, 2, \dots, n$), co oznacza, że niewielkim zmianom prawdopodobieństw powinna odpowiadać niewielka zmiana entropii.

P2. Jeżeli wszystkie n zdarzeń zmiennej losowej X są jednakowo prawdopodobne, to funkcja $H_S(p(x_1), p(x_2), \dots, p(x_n))$ powinna rosnać monotonicznie wraz ze wzrostem n , gdzie $p(x_1) = p(x_2) = \dots = p(x_n) = \frac{1}{n}$.

P3. Funkcja $H_S(p(x_1), p(x_2), \dots, p(x_n))$ powinna być symetryczna, co oznacza, że wartość funkcji entropii jest niezmiennikiem permutacji prawdopodobieństw $p(x_1), p(x_2), \dots, p(x_n)$.

P4. Funkcja $H_S(p(x_1), p(x_2), \dots, p(x_n))$ powinna być koherentna, co oznacza w tym przypadku, że jeżeli realizacja zdarzeń odbywa się w dwóch kolejno następujących po sobie etapach, to entropia początkowa powinna być sumą ważoną entropii poszczególnych etapów.

Istnieje dokładnie jedna, z dokładnością do stałej k [4], funkcja $H_S(p(x_1), p(x_2), \dots, p(x_n))$ n -zmiennych spełniająca powyższe warunki i jest ona określona wzorem:

$$H_S(X) = H_S(p(x_1), p(x_2), \dots, p(x_n)) = k \sum_{i=1}^n p(x_i) \log \frac{1}{p(x_i)}. \quad (1)$$

Dowód tego twierdzenia przedstawiono między innymi w pracach ([4], [12]).

Stała k we wzorze (1) stanowi wybór jednostki entropii. Jeżeli $k = \frac{1}{\log 2}$ jednostką entropii jest bit, a funkcja zapisana wzorem (1) przyjmuje postać:

$$H_S(X) = H_S(p(x_1), p(x_2), \dots, p(x_n)) = \sum_{i=1}^n p(x_i) \log_2 \frac{1}{p(x_i)}. \quad (2)$$

Entropia Shannona H_S zmiennej losowej X , której wartości występują z prawdopodobieństwami $\{p(x_1), p(x_2), \dots, p(x_n)\}$ jest określona wzorem (2) i jest traktowana jako miara niepewności związanej z rozkładem dyskretnym o prawdopodobieństwach $\{p(x_1), p(x_2), \dots, p(x_n)\}$ [25].

Określenie shannonowskiej entropii pozwala ustalić szereg jej istotnych własności, przedstawionych między innymi w pracy [20]:

- 1) Entropia Shannona jest wielkością nieujemną, $\forall p(x_i) \in [0, 1] \ H_S(X) \geq 0$.
- 2) Entropia H_S przyjmuje wartość 0, gdy $p(x_i) = 1$ dla pewnego i ($i = 1, 2, \dots, n$).
- 3) Entropia H_S przyjmuje wartość największą równą $H_S(X) = \log_2 n$, gdy wszystkie prawdopodobieństwa $p(x_i)$ są sobie równe dla $i = 1, 2, \dots, n$.
- 4) Entropia H_S jest wklęsła.

5) Entropia Shannona spełnia własność addytywności dla dyskretnych zmiennych losowych niezależnych.

Własność 5 dotyczy pary zmiennych losowych niezależnych. Jeśli zmienne losowe dyskretne X i Y o rozkładach prawdopodobieństwa odpowiednio $\{p(x_i)\}$ oraz $\{q(y_j)\}$ dla $i = 1, 2, \dots, n$ oraz $j = 1, 2, \dots, m$ są niezależne, to rozkład prawdopodobieństwa zmiennej losowej dwuwymiarowej (X, Y) zadany jest iloczynem $\{p(x_i)q(y_j)\}$,

gdzie $\sum_{i=1}^n \sum_{j=1}^m p(x_i)q(y_j) = 1$. Wtedy entropia Shannona spełnia własność addytywności,

a zatem prawdziwa jest równość $H_S(X, Y) = H_S(X) + H_S(Y)$. Prawdziwość podanej równości wynika z własności funkcji logarytmicznych [20].

3. ENTROPIA RÉNYIEGO

Węgierski matematyk Alfred Rényi wśród swoich licznych osiągnięć z dziedziny teorii prawdopodobieństwa zaproponował uogólnienie pojęcia entropii Shannona. Wyprowadzenie uogólnionej postaci entropii wymaga zastosowania średniej Kołmogorowa-Nagumo (KN) [15].

Niech $\varphi(x_i)$ będzie funkcją ciągłą i ściśle monotoniczną. Średnia Kołmogorowa-Nagumo związana z funkcją $\varphi(x_i)$ zmiennej losowej dyskretnej X przyjmującej wartości $\{x_1, x_2, \dots, x_n\}$, definiowana jest następująco [13]:

$$\langle X \rangle_\varphi = \varphi^{-1} \left(\sum_{i=1}^n p(x_i) \varphi(x_i) \right). \quad (3)$$

Średnia arytmetyczna jest średnią Kołmogorowa-Nagumo związaną z funkcją liniową $\varphi(x_i) = x_i$, uwzględniając bowiem w formule (3) zamiast $\varphi(x_i)$ wartość x_i uzyskana zostanie znana postać wartości oczekiwanej.

Średnie Kołmogorowa-Nagumo są niezmiennicze ze względu na transformacje afiniczne $\varphi(x_i) \rightarrow \gamma(x_i) = a\varphi(x_i) + b$, gdzie a, b są liczbami. Stąd funkcją odwrotną do $\varphi(x_i)$ jest $\gamma^{-1}(x_i) = \varphi^{-1} \left(\frac{x_i - b}{a} \right)$, a średnia KN związana w funkcją $\varphi(x_i)$ zmiennej losowej X , czyli $\langle X \rangle_\varphi$ równa jest średniej KN $\langle X \rangle_\gamma$ związanej z funkcją $\gamma(x_i)$.

Rozważmy dwie niezależne zmienne losowe dyskretne X i Y o rozkładach prawdopodobieństwa odpowiednio $\{p(x_i)\}$ oraz $\{q(y_i)\}$ dla $i = 1, 2, \dots, n$. Poszukiwanie funkcji $\varphi(x_i)$ spełniającej własność addytywności średnich $\langle X + Y \rangle_\varphi = \langle X \rangle_\varphi + \langle Y \rangle_\varphi$ daje, oprócz funkcji typu $ax_i + b$, jedynie funkcję postaci $\varphi(x_i) = a\mu^{vx_i} + b$, gdzie a, b są liczbami rzeczywistymi. Warunek addytywności spełniają jedynie średnie KN związane z dwiema funkcjami, pierwsza z nich to funkcja liniowa $\varphi(x_i) = x_i$, druga to funkcja wykładnicza $\varphi(x_i) = a\mu^{vx_i} + b$ [15]. Rozpatrzmy dalej funkcję $\varphi(x_i) = \mu^{vx_i}$ ($a = 1, b = 0$),

przyjmując $\mu = 2$, $\nu = 1 - \alpha$. Wtedy funkcja $\varphi(x_i)$ przyjmuje postać $2^{(1-\alpha)x_i}$, a funkcja odwrotna $\varphi^{-1}(x_i) = \frac{1}{1-\alpha} \log_2 x_i$.

Średnia ilość informacji w sensie Kołmogorowa-Nagumo związana z funkcją $\varphi(x_i) = 2^{(1-\alpha)x_i}$ i rozkładem prawdopodobieństwa $\{p(x_i)\}$ zmiennej losowej X jest entropią Rényiego. W teorii informacji ilość informacji dostarczaną przez wiadomość o przyjęciu przez dyskretną zmienną losową X wartości x_i wyraża się miarą $I(x_i) = -\log_2 p(x_i)$ [25]. Stąd uwzględniając przyjęte powyżej założenia, jawną postać entropii Rényiego otrzymuje się drogą poniższych przekształceń ([5], [13]):

$$\begin{aligned} H_{R\alpha}(X) &= \varphi^{-1} \left(\sum_{i=1}^n p(x_i) \varphi(I(x_i)) \right) \\ &= \varphi^{-1} \left(\sum_{i=1}^n p(x_i) \varphi(-\log_2 p(x_i)) \right). \end{aligned} \quad (4)$$

Uwzględnivszy, że przyjęto powyżej funkcję $\varphi(x_i)$ postaci $2^{(1-\alpha)x_i}$ wraz z funkcją do niej odwrotną $\varphi^{-1}(x_i) = \frac{1}{1-\alpha} \log_2 x_i$, wyrażenie (4) można wyrazić następująco:

$$\begin{aligned} &\frac{1}{1-\alpha} \log_2 \left(\sum_{i=1}^n p(x_i) \cdot 2^{-(1-\alpha)\log_2 p(x_i)} \right) \\ &= \frac{1}{1-\alpha} \log_2 \left(\sum_{i=1}^n p(x_i) \cdot 2^{\log_2 p(x_i)(\alpha-1)} \right) \\ &= \frac{1}{1-\alpha} \log_2 \left(\sum_{i=1}^n p(x_i) p(x_i)^{(\alpha-1)} \right) \\ &= \frac{1}{1-\alpha} \log_2 \left(\sum_{i=1}^n p(x_i)^\alpha \right). \end{aligned} \quad (5)$$

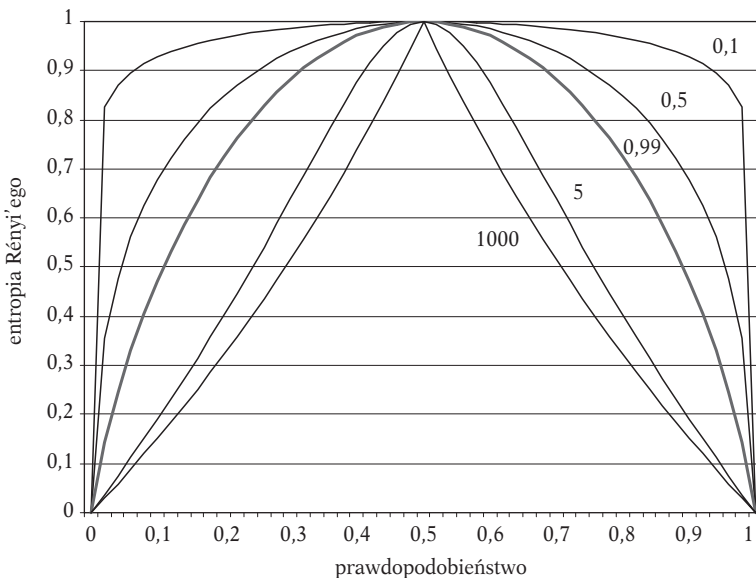
Postać $\frac{1}{1-\alpha} \log_2 \left(\sum_{i=1}^n p(x_i)^\alpha \right)$ powyższej formuły jest podawana najczęściej w literaturze jako definicja entropii Rényiego, a jej związek ze średnią Kołmogorowa-Nagumo jest mniej znany. Zwracają na to uwagę Czachor i Naudts w pracy [5] oraz Lavenda [13]. Entropia $H_{R\alpha}(X)$ określona wzorem (5) nazywana jest najczęściej entropią Rényiego stopnia α lub α -entropią Rényiego dla $\alpha > 0$, $\alpha \neq 1$.

Entropia Rényiego stopnia α dyskretnej zmiennej losowej X o rozkładzie prawdopodobieństwa $\{p(x_i)\}$ $\left(0 \leq p(x_i) \leq 1, \sum_{i=1}^n p(x_i) = 1 \right)$ spełnia następujące własności:

- I. Jest wielkością nieujemną $\forall \alpha \in (0, 1) \cup (1, +\infty)$, $H_{R\alpha}(X) \geq 0$.
- II. Jest wklęsła $\forall \alpha \in (0, 1)$ [8].
- III. Jest wklęsła lub wypukła dla $\alpha > 1$ [8].
- IV. Przyjmuje wartość 0, gdy $p(x_i) = 1$ dla pewnego i ($i = 1, 2, \dots, n$).
- V. Przyjmuje wartość największą równą $H_{R\alpha}(X) = \log_2 n$, gdy wszystkie prawdopodobieństwa $p(x_i)$ są sobie równe dla $i = 1, 2, \dots, n$.
- VI. Entropia Rényiego spełnia własność addytywności dla zmiennych losowych niezależnych.
- VII. Entropie Shannona H_S oraz Rényiego $H_{R\alpha}$ spełniają relację $H_{R\alpha_1} \geq H_S \geq H_{R\alpha_2}$ jeżeli $0 < \alpha_1 < 1$ oraz $\alpha_2 > 1$ [19].
- VIII. Entropia Shannona H_S jest granicą entropii Rényiego $H_{R\alpha}$ dla $\alpha \rightarrow 1$,

$$\lim_{\alpha \rightarrow 1} \frac{1}{1 - \alpha} \log_2 \left(\sum_{i=1}^n p(x_i)^\alpha \right) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i).$$

Z własności V wynika, że dla zmiennej losowej o rozkładzie równomiernym ($p(x_1) = p(x_2) = \dots = p(x_n)$) wartość największa entropii Rényiego równa jest entropii Shannona dla tego rozkładu. Zatem w nierówności $H_{R\alpha_1} \geq H_S \geq H_{R\alpha_2}$ podanej we własności VII zachodzi równość dla rozkładu równomiernego. Wartości entropii Rényiego rosną monotonicznie wraz ze wzrostem n , co jest spełnieniem jednego z postulatów stawianych funkcji entropii przez Shannona.



Rysunek 1. Entropia Rényiego zmiennej losowej o rozkładzie prawdopodobieństwa $\{p, (1 - p)\}$

Własność VI dotyczy pary zmiennych losowych niezależnych. Jeśli zmienne losowe dyskretne X i Y o rozkładach prawdopodobieństwa odpowiednio $\{p(x_i)\}$ oraz $\{q(y_j)\}$ dla $i = 1, 2, \dots, n$ oraz $j = 1, 2, \dots, m$ są niezależne, prawdziwa jest równość $H_{R\alpha}(X, Y) = H_{R\alpha}(X) + H_{R\alpha}(Y)$. Prawdziwość podanej równości wynika, podobnie jak w przypadku równości dla entropii Shannona, z własności funkcji logarytmicznych.

Entropia Rényiego dla zmiennej losowej o dwumianowym rozkładzie prawdopodobieństwa $\{p, (1-p)\}$ przyjmuje wartość największą równą jeden dla $p = 1-p$, czyli dla rozkładu o prawdopodobieństwach $\{0,5; 0,5\}$. Na rysunku 1 przedstawiono wykres entropii Rényiego dla wybranych wartości stopnia α rozkładu o prawdopodobieństwach $\{p, (1-p)\}$. Dla $\alpha = 0,99$ wykres entropii Rényiego jest przybliżony do wykresu entropii Shannona.

4. ENTROPIA TSALLISA

Entropię określaną mianem „entropii typu α ” zaproponowali niezależnie w swoich pracach z roku 1970 Havrda i Charvát oraz Daróczy, a następnie Tsallis w roku 1988 przy okazji tzw. termodynamiki nieekstensywnej [5]. Stąd często w literaturze pojawia się określenie entropia Havrda-Charvát-Daróczy-Tsallis (HCDT) lub w pracach z zakresu fizyki, entropia Tsallisa. Entropia HCDT została zaproponowana jako pierwsza nie-logarytmiczna postać entropii.

Do wyprowadzenia ogólnej postaci entropii HCDT wykorzystane zostaną ponownie własności średniej Kołmogorowa-Nagumo dotyczące niezmienniczości ze względu na transformacje afiniczne $\varphi(x_i) \rightarrow \gamma(x_i) = a\varphi(x_i) + b$. Z rodziny funkcji typu

$$\varphi(x_i) = a\mu^{vx_i} + b \text{ wybrana zostanie funkcja } \varphi(x_i) = \frac{1}{1-\alpha} 2^{(1-\alpha)x_i} - \frac{1}{1-\alpha} = \frac{2^{(1-\alpha)x_i} - 1}{1-\alpha},$$

gdzie $\alpha \in (0, 1) \cup (1, +\infty)$.

Postać entropii HCDT uzyskana jest drogą modyfikacji średniej Kołmogorowa-Nagumo związanej z funkcją $\varphi(x_i) = \frac{1}{1-\alpha} 2^{(1-\alpha)x_i} - \frac{1}{1-\alpha}$:

$$\begin{aligned} H_{R\alpha} &= \varphi^{-1} \left(\sum_{i=1}^n p(x_i) \varphi(-\log_2 p(x_i)) \right) \\ &= \varphi^{-1} \left(\sum_{i=1}^n p(x_i) \frac{2^{(1-\alpha)(-\log_2 p(x_i))} - 1}{1-\alpha} \right) \\ &= \varphi^{-1} \left(\sum_{i=1}^n p(x_i) \frac{2^{\log_2 \left(\frac{1}{p(x_i)} \right)^{1-\alpha}} - 1}{1-\alpha} \right). \end{aligned} \tag{6}$$

Wykorzystując własności funkcji wykładniczych i logarytmicznych wyrażenie $2^{\log_2 \left(\frac{1}{p(x_i)} \right)^{1-\alpha}}$ można wyrazić w prostszej postaci $\left(\frac{1}{p(x_i)} \right)^{1-\alpha}$. Stąd wyrażenie (6) przyjmuje postać:

$$\begin{aligned} & \varphi^{-1} \left(\frac{\sum_{i=1}^n p(x_i) \left(\frac{1}{p(x_i)} \right)^{1-\alpha} - 1}{1 - \alpha} \right) \\ &= \varphi^{-1} \left(\frac{\sum_{i=1}^n p(x_i)^\alpha - 1}{1 - \alpha} \right). \end{aligned} \quad (7)$$

Wyrażenie uzyskane w drodze przekształceń prowadzących do otrzymania entropii $H_{T\alpha}(X)$ postaci:

$$H_{T\alpha}(X) = \frac{\sum_{i=1}^n p(x_i)^\alpha - 1}{1 - \alpha} \quad (8)$$

nosi miano entropii Tsallisa lub entropii Havrda-Charvát-Daróczy-Tsallisa.

Entropia Tsallisa typu α dyskretnej zmiennej losowej X o rozkładzie prawdopodobieństwa $\{p(x_i)\}$ $\left(0 \leq p(x_i) \leq 1, \sum_{i=1}^n p(x_i) = 1 \right)$ spełnia następujące własności:

- Jest wielkością nieujemną $\forall \alpha \in (0, 1) \cup (1, +\infty)$, $H_{T\alpha}(X) \geq 0$.
- Jest wklęsła $\forall \alpha \in (0, 1) \cup (1, +\infty)$
- Przyjmuje wartość 0, gdy $p(x_i) = 1$ dla pewnego i ($i = 1, 2, \dots, n$).
- Przyjmuje wartość największą, gdy wszystkie prawdopodobieństwa $p(x_i)$ są sobie równe dla $i = 1, 2, \dots, n$.
- Dla $\forall \alpha_1, \alpha_2 \in (0, 1) \cup (1, +\infty)$ jeżeli $\alpha_1 < \alpha_2$ zachodzi relacja $H_{T\alpha_1}(X) > H_{T\alpha_2}(X)$.
- Iloczyn entropii Shannona H_S i $\ln 2$ jest granicą entropii Tsallisa $H_{T\alpha}$ dla $\alpha \rightarrow 1$;

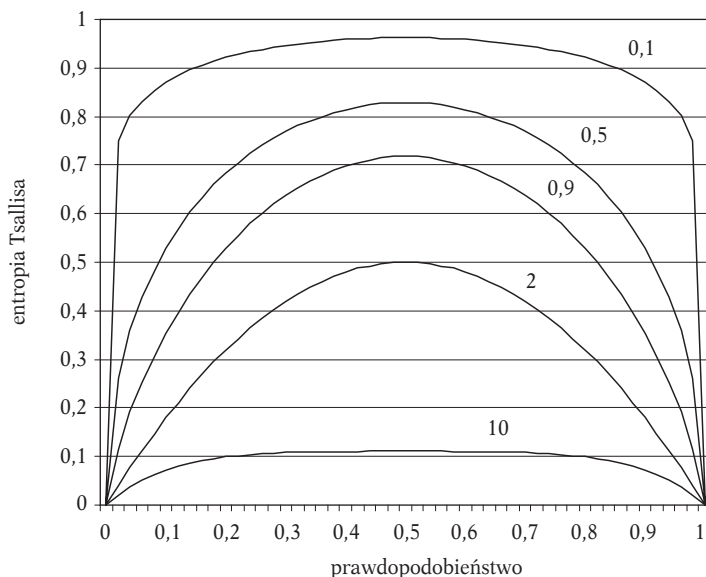
$$\lim_{\alpha \rightarrow 1} \frac{\sum_{i=1}^n p(x_i)^\alpha - 1}{1 - \alpha} = -\ln 2 \sum_{i=1}^n p(x_i) \log p(x_i) = \ln 2 H_S(X).$$

- Spełnia tzw. własność pseudo-addytywności (subaddytywności) dla zmiennych losowych niezależnych [13]:

$$H_{T\alpha}(X, Y) = H_{T\alpha}(X) + H_{T\alpha}(Y) + (1 - \alpha) H_{T\alpha}(X) H_{T\alpha}(Y).$$

Z własności pseudo-addytywności wynika, że dla $\alpha \in (0,1)$ entropia $H_{T\alpha}(X, Y)$ jest większa od sumy $H_{T\alpha}(X) + H_{T\alpha}(Y)$, natomiast dla $\alpha \in (1, +\infty)$ zachodzi relacja $H_{T\alpha}(X, Y) < H_{T\alpha}(X) + H_{T\alpha}(Y)$. Własność pseudo-addytywności w fizyce nazywana jest często nieekstensywnością. Własność pseudo-addytywności odróżnia entropię Tsallisa od entropii Shannona i Rényiego, które spełniały własność addytywności dla dyskretnych zmiennych losowych niezależnych X i Y .

Entropia HCDT przyjmuje wartość największą dla zmiennej losowej o rozkładzie $\{p, (1-p)\}$ dla $p = 0,5$. Wartość ta jest funkcją stopnia α oraz jest mniejsza dla każdego $\alpha \in (0,1) \cup (1, +\infty)$ od wartości maksymalnej entropii Shannona i Rényiego dla tego rozkładu $\{0,5; 0,5\}$. Na rysunku 2 przedstawiono wykres entropii Tsallisa dla wybranych wartości stopnia α rozkładu o prawdopodobieństwach $\{p, (1-p)\}$.



Rysunek 2. Entropia Tsallisa zmiennej losowej o rozkładzie prawdopodobieństwa $\{p, (1-p)\}$

Źródło: opracowanie własne.

Entropia Tsallisa osiąga wartość największą dla równomiernego rozkładu prawdopodobieństwa ($p(x_1) = p(x_2) = \dots = p(x_n)$). Wartość największa entropii HCDT jest funkcją nie tylko wartości n , jak to było w przypadku entropii Shannona oraz Rényiego, ale i stopnia α . Wraz ze wzrostem n rośnie wartość entropii $H_{T\alpha}$ dla danego stopnia α . Wartości największe omawianych w artykule entropii zamieszczono w tabeli 1.

Tabela 1

Wartości entropii rozkładu równomiernego zmiennej losowej dyskretnej

n	Entropia					
	Shannona	Rényiego $\alpha \in (0,1) \cup (1, +\infty)$	Tsallisa			
			$\alpha = 0,1$	$\alpha = 0,5$	$\alpha = 2$	$\alpha = 10$
2	1,00000	1,00000	0,96230	0,82843	0,50000	0,11089
4	2,00000	2,00000	2,75800	2,00000	0,75000	0,11111
5	2,32193	2,32193	3,61856	2,47214	0,80000	0,11111
10	3,32193	3,32193	7,71476	4,32456	0,90000	0,11111
50	5,64386	5,64386	36,45796	12,14214	0,98000	0,11111
100	6,64386	6,64386	68,99526	18,00000	0,99000	0,11111
200	7,64386	7,64386	129,71200	26,28427	0,99500	0,11111

Źródło: opracowanie własne.

5. MIARY KONCENTRACJI BAZUJĄCE NA ENTROPII

Związek pomiędzy miarami koncentracji, czy też miarami dekoncentracji, a entropią wydaje się być oczywisty. Entropia jest bowiem miarą nieokreśloności i niejednorodności rozkładu dyskretnej zmiennej losowej X o rozkładzie prawdopodobieństwa $\{p(x_1), p(x_2), \dots, p(x_n)\}$ z prawdopodobieństwami $\{p(x_i)\}$ spełniającymi warunek

$$0 \leq p(x_i) \leq 1, \sum_{i=1}^n p(x_i) = 1. \text{ Jednym z przykładów związku miar koncentracji z entropią}$$

jest indeks Giniego-Simpsona. Indeks ten jest miarą równomierności dyskretnego rozkładu prawdopodobieństwa i określony jest formułą [17]:

$$GS = 1 - \sum_{i=1}^n p(x_i)^2. \quad (9)$$

Miernik ten jest szczególnym przypadkiem entropii Tsallisa $H_{T\alpha}(X)$. Jeśli bowiem w formule (8) uwzględniony zostanie stopień α równy dwa, entropia HCDT przyjmie następującą postać:

$$H_{T2}(X) = \frac{1}{\alpha - 1} \left[1 - \sum_{i=1}^n p(x_i)^\alpha \right] = \left[1 - \sum_{i=1}^n p(x_i)^2 \right] = GS.$$

Miernik ten nazywany jest także w literaturze entropią kwadratową Vajdy (zob. [16]).

Warto zauważyć, że suma kwadratów prawdopodobieństw $\sum_{i=1}^n p(x_i)^2$ występująca w formule indeksu Giniego-Simpsona jest znanym w literaturze miernikiem koncentracji, zwanym indeksem Herfindahla-Hirschmana (*HHI*).

Konstrukcja indeksu Herfindahla-Hirschmana wskazuje, że wybór wykładnika potęgi prawdopodobieństw $p(x_i)$ implikuje relatywny wpływ na pomiar koncentracji wartości x_i , zachodzących odpowiednio z małymi bądź znacznymi prawdopodobieństwami. Dobór wyższych wykładników potęgi sprawia, że większe znaczenie przypisywane jest tym wartościom zmiennej losowej X , które występują z dużym prawdopodobieństwem. Stąd w literaturze pojawiły się modyfikacje i uogólnienia indeksu *HHI*. Hannah i Kay zaproponowali w 1977 roku swobodny dobór wykładnika, podając następującą formułę miernika koncentracji [2]:

$$HKI_{(\alpha)} = \begin{cases} \left[\sum_{i=1}^n p(x_i)^\alpha \right]^{\frac{1}{\alpha-1}} & \text{dla } \alpha > 0, \alpha \neq 1 \\ \exp \left[\sum_{i=1}^n p(x_i) \ln p(x_i) \right] & \text{dla } \alpha = 1 \end{cases} \quad (10)$$

Indeks Hannah-Kaya dla $\alpha = 1$ ($HKI_{(1)}$) jest granicą indeksu $HKI_{(\alpha)}$ dla $\alpha \rightarrow 1$, zaś w przypadku przyjęcia w formule (10) $\alpha = 2$ indeks Hannah-Kaya staje się indeksem Herfindahla-Hirschmana. Związki indeksu Hannah-Kaya z miarami entropii określają Bajo i Salas (2002) w pracy [2].

Odrębną klasę mierników koncentracji stanowią miary wykorzystujące definicję entropii Shannona. Entropia Shannona $H_S(X)$ zdefiniowana formułą (2) jest odwrotnością stopnia koncentracji prawdopodobieństw $p(x_i)$ w rozkładzie prawdopodobieństwa $\{p(x_1), p(x_2), \dots, p(x_n)\}$. Entropia $H_S(X)$ równa jest bowiem zero, gdy jedna z wartości x_i ($i = 1, 2, \dots, n$) osiągana jest z prawdopodobieństwem równym jedności. W przypadku gdy wszystkie n wartości dyskretnej zmiennej losowej X zachodzą z jednakowymi prawdopodobieństwami, entropia $H_S(X)$ osiąga wartość największą równą $H_S^{\max}(X) = \log_2 n$. W wielu pracach entropia Shannona wykorzystywana jest bezpośrednio do oceny stopnia niejednorodności rozkładu badanej cechy [3] bądź do analizy stopnia dywersyfikacji działalności przedsiębiorstw na poziomie korporacyjnym (zob. [9], [10]) czy też na poziomie regionalnym (zob. [1], [7]).

Znacznie bardziej popularne w naukach ekonomicznych są mierniki będące funkcją wystandaryzowanej postaci entropii. Mierniki te stanowią miarę koncentracji, tak jak miara

$$h = 1 - \frac{H_S(X)}{H_S^{\max}(X)} \quad (11)$$

przedstawiona przez Roeske-Słomka (1994) w pracy [21] bądź też miarę dekoncentracji postaci:

$$e = \frac{H_S(X)}{H_S^{\max}(X)} \quad (12)$$

proponowaną między innymi w pracach ([17], [23], [24]).

Z konstrukcji miary (11) wynika, że w pomiarze stopnia koncentracji dyskretnego rozkładu prawdopodobieństwa największy wpływ na otrzymane wyniki mają wartości zmiennej losowej X realizowane z dużym prawdopodobieństwem. Powoduje to sytuację, że zdecydowana większość obszaru zmienności miary h przeznaczona jest dla rozkładów bliskich pełnej koncentracji.

W niniejszym artykule skonstruowana zostanie miara koncentracji dyskretnego rozkładu prawdopodobieństwa $\{p(x_1), p(x_2), \dots, p(x_n)\}$ bazująca na entropii Rényiego $H_{R\alpha}(X)$. Proponowana miara jest jednocześnie uogólnieniem miary (11) i zdefiniowana zostanie następującą formułą:

$$HRI_{\alpha}(X) = \begin{cases} 1 - \frac{\log_2 \left(\sum_{i=1}^n p(x_i)^{\alpha} \right)}{(1 - \alpha) \log_2 n}, & \text{dla } \alpha > 0 \text{ i } \alpha \neq 1 \\ 1 - \frac{\sum_{i=1}^n p(x_i) \log_2 \frac{1}{p(x_i)}}{\log_2 n}, & \text{dla } \alpha = 1 \end{cases} \quad (13)$$

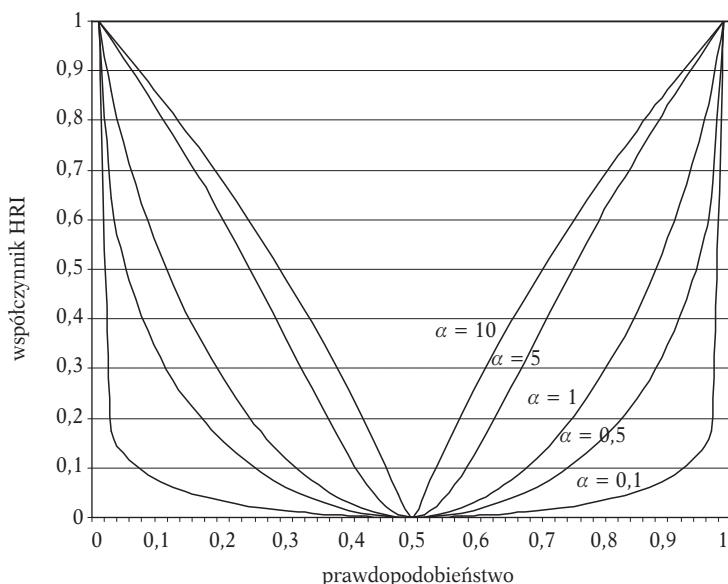
Widać, że proponowaną miarę można następująco wyrazić jako funkcję entropii Rényiego:

$$HRI_{\alpha}(X) = \begin{cases} 1 - \frac{H_{R\alpha}(X)}{H_{R\alpha}^{\max}(X)}, & \text{dla } \alpha > 0 \text{ i } \alpha \neq 1 \\ 1 - \frac{H_S(X)}{H_S^{\max}(X)}, & \text{dla } \alpha = 1 \end{cases} \quad (14)$$

Własności entropii Rényiego determinują własności miary HRI_{α} . Miara koncentracji HRI_{α} jest symetryczna, to oznacza, że jest niezależna od permutacji prawdopodobieństw $p(x_i)$ ($i = 1, 2, \dots, n$) w dyskretnym rozkładzie prawdopodobieństwa $\{p(x_1), p(x_2), \dots, p(x_n)\}$. Spełnia także kryterium unormowania w przedziale $[0,1]$ i jest jednoznacznie określona dla krańcowych sytuacji. W przypadku równomiernego rozkładu prawdopodobieństw $p(x_i)$ miernik HRI_{α} przyjmuje wartość zero, natomiast w sytuacji pełnej koncentracji osiąga wartość jeden. Uogólnienie miary (11) bazujące na entropii Shannona daje możliwość swobodnego wyboru potęgi prawdopodobieństw $p(x_i)$. Dobór stopnia α entropii Rényiego $H_{R\alpha}(X)$ umożliwia relatywny wpływ na ocenę stopnia koncentracji wartości x_i zmiennej losowej X występujących z odpowiednio

niskimi bądź wysokimi prawdopodobieństwami. Dla $\alpha \in (0,1]$ zdecydowanie większa część przedziału zmienności miary HRI_α przeznaczona jest dla rozkładów prawdopodobieństw bliskich pełnej koncentracji. Występuje znikoma reakcja miernika przy odbieganiu rozkładu $\{p(x_1), p(x_2), \dots, p(x_n)\}$ od rozkładu równomiernego o prawdopodobieństwach $p(x_1) = p(x_2) = \dots = p(x_n) = \frac{1}{n}$. Dla rozkładów bliskich pełnej koncentracji miernik HRI_α przyjmuje wartości z dolnego krańca przedziału $[0,1]$. Zatem omawiany miernik dla stopnia α z przedziału $(0,1]$ charakteryzuje się niewielką awersją do niejednorodnego rozkładu prawdopodobieństw.

Proponowana miara HRI_α jest wypukła dla każdego stopnia $\alpha \in (0,1]$. Na rysunku 3 przedstawiono wartości miary HRI_α dla rozkładu o prawdopodobieństwach $\{p, (1-p)\}$, przyjmując wybrane wartości stopnia α .



Rysunek 3. Wartości miary HRI_α dla zmiennej losowej o rozkładzie prawdopodobieństwa $\{p, (1-p)\}$

Źródło: opracowanie własne.

Miara koncentracji HRI_α dla α większego od jedności jest wypukła bądź wklęsła. Charakteryzuje się większą wrażliwością (bezwzględną i względną) na zmiany w rozkładzie prawdopodobieństwa $\{p(x_1), p(x_2), \dots, p(x_n)\}$. Wraz ze wzrostem α zwiększa się awersja miernika do nierównomiernego rozkładu prawdopodobieństw, z jakim zachodzą wartości x_i dyskretnej zmiennej losowej X .

6. PODSUMOWANIE

Możliwość wykorzystania entropii do określania stopnia nieokreśloności badanych stanów czy też rozkładów prawdopodobieństw przyczyniło się do określania różnych postaci entropii. Zastosowanie pojęcia entropii w teorii informacji przez Shannona dało początek licznym uogólnieniom miary entropii. W artykule wskazano jedynie wybrane uogólnienia entropii Shannona, mianowicie entropię Rényiego oraz Tsallisa. Entropia Rényiego stopnia α doczekała się kolejnych uogólnień, czego przykładem jest chociażby entropia stopni (α, β) zaproponowana przez Kapura (1967). Inne probabilistyczne miary entropii proponują między innymi Lavenda [14], Sharma i Taneja [22], Chakrabarti [4], Garbaczewski [6], Nanda i Paul [18], Kumar i in. [11].

Entropia Shannona jest granicą entropii Rényiego oraz Tsallisa dla stopnia α dążącego do jedności. Entropie Shannona i Rényiego mają postać funkcji logarytmicznych, natomiast entropia Tsallisa jest pierwszą proponowaną w literaturze nie-logarytmiczną entropią. Warto podkreślić, że entropie Shannona i Rényiego są addytywne dla dyskretnych zmiennych losowych X i Y , natomiast entropia Tsallisa nie spełnia własności addytywności, a jedynie tzw. własność pseudo-addytywności.

Wartości entropii Shannona, Rényiego oraz Tsallisa zależą jedynie od prawdopodobieństw, jakie towarzyszą realizacji wartości $\{x_1, x_2, \dots, x_n\}$ dyskretnej zmiennej losowej X , a nie od tych wartości. Wartość największą każda ze wskazanych entropii osiąga dla rozkładu równomiernego o prawdopodobieństwach $p(x_1) = p(x_2) = \dots = p(x_n) = \frac{1}{n}$, przy czym monotoniczny wzrost entropii wraz ze wzrostem n . Entropia jest odwrotnością stopnia koncentracji prawdopodobieństw $p(x_i)$ w dyskretnym rozkładzie prawdopodobieństwa $\{p(x_1), p(x_2), \dots, p(x_n)\}$. Stąd znajomość entropii może być przydatna do rozpoznania stopnia koncentracji rozkładu prawdopodobieństwa zmiennej losowej dyskretnej. W artykule zaproponowano miarę umożliwiającą ocenę stopnia koncentracji dyskretnego rozkładu prawdopodobieństwa będącą funkcją entropii Rényiego. Istotną zaletą proponowanej miary jest swobodny dobór potęgi prawdopodobieństw $p(x_i)$. Dobór stopnia α umożliwia bowiem relatywny wpływ na ocenę stopnia koncentracji rozkładu wartości x_i zmiennej losowej X występujących z odpowiednio niskimi bądź wysokimi prawdopodobieństwami.

Uniwersytet Warmińsko-Mazurski w Olsztynie

LITERATURA

- [1] Attaran M., [1985], *Industrial diversity and economic performance in U.S. areas*, The Annals of Regional Science 20, s. 44-54.
- [2] Bajo O., Salas R., [2002], *Inequality foundations of concentration measures: An application to the Hannah-Kay indices*, Spanish Economic Review 4, s. 311-316.
- [3] Campos D., Isaza J.F., [2009], *A geometrical index for measuring species diversity*, Ecological Indicators 9, s. 651-658.
- [4] Chakrabarti C.G., Chakrabarty I., [2005], *Shannon entropy: axiomatic characterization and application*, International Journal of Mathematics and Mathematical Sciences, Vol. 17, s. 2847-2854.

- [5] Czachor M., Naudts J., [2002], *Thermostatistics based on Kolmogorov-Nagumo averages: unifying framework for extensive and nonextensive generalizations*, Physics Letters A 298, s. 369-374.
- [6] Garbaczewski P., [2006], *Differential Entropy and Dynamics of Uncertainty*, „Journal of Statistical Physics”, Vol. 123, No. 2, s. 315-355.
- [7] Hackbart M.W., Anderson D.A., [1975], *On measuring economic diversification*, Land Economics 51, s. 374-378.
- [8] Hibbard L.S., [2004], *Region segmentation using information divergence measures*, Medical Image Analysis 8, s. 233-244.
- [9] Jacquemin A.P., Berry C.H., [1979], *Entropy measure of diversification and corporate growth*, „Journal of Industrial Economics” 27 (4), s. 359-369.
- [10] Koen F., [2007], *Entropy and information theory*, [w:] Hanusch H., Pyka A. (red.), The Elgar Companion to Neo-Schumpeterian Economics.
- [11] Kumar R., Kumar S., Kumar A., [2010], *A New Measure of Probabilistic Entropy and its Properties*, Applied Mathematical Sciences, Vol. 4, s. 1387-1394.
- [12] Kuriata E., [2001], *Teoria informacji i kodowania*, Oficyna Wydawnicza Politechniki Zielonogórskiej.
- [13] Lavenda B.H., [2004], *Mean Entropies*, Open Sys. Information Dyn. 12, s. 289-302.
- [14] Lavenda B.H., [2006], *Geometric Entropies of Mixing (EOM)*, Open Sys. Information Dyn. 13, s. 91-101.
- [15] Masi M., [2005], *A step beyond Tsallis and Rényi entropies*, Physics Letters A 338, s. 217-224.
- [16] Mayoral M.M., [1998], *Rényi's entropy as an index of diversity in simple-stage cluster sampling*, „Journal of Information Sciences” 105, s. 101-114.
- [17] McBratney A., Minasny B., [2007], *On measuring pedodiversity*, Geoderma 141, s. 149-154.
- [18] Nanda A.K., Paul P., [2006], *Some results on generalized residual entropy*, Information Sciences 176 (1), s. 27-47.
- [19] Principe J.C., Xu D., Fisher III J.W., [????], *Information-theoretic learning*, in: Haykin S. (ed.), Unsupervised Adaptive Filtering. Wiley, New York, s. 265-319.
- [20] Przybyszewski R., Wędrawska E., [2005], *Algorytmiczna teoria entropii*, „Przegląd Statystyczny” nr 2, tom 52 Warszawa, s. 85-102.
- [21] Roeske-Słomka I., [1994], *Klasyczne a entropijne miary koncentracji i asymetrii w badaniu rozmieszczenia ludności*, Zesz. Nauk. AE Pozn. Seria I, nr 219, s. 48-59.
- [22] Sharma B.D., Taneja I.J., [1975], *Entropy of type (α, β) and other generalized measures in information theory*, Metrika 22, s. 205-215.
- [23] Troutt M. D., Acar W., [2005], *A Lorenz-Pareto measures of pure diversification*, „European Journal of Operational Research” 167, s. 543-549.
- [24] Wędrawska E., [2003], *Datalogiczna miara ilości informacji strukturalnej jako instrument zarządzania zasobami informacji statystycznej*, [w:] „Pozyskiwanie wiedzy i zarządzanie wiedzą”, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, nr 975 Wrocław, s. 447-459.
- [25] Woodward P.M., [1959], *Wstęp do teorii informacji*, PWN, Warszawa.

Praca wpłynęła do redakcji kwiecień 2010 r.

WYKORZYSTANIE ENTROPII SHANNONA I JEJ UOGÓLNIENÍ DO BADANIA ROZKŁADU PRAWDOPODOBIEŃSTWA ZMIENNEJ LOSOWEJ DYSKRETNEJ

Streszczenie

O entropii można mówić wszędzie tam, gdzie istnieje potrzeba rozpoznania i zmierzenia nieokreśloności badanych rozkładów zmiennych losowych. Potrzeba zmierzenia stopnia koncentracji występuje w badaniu wielu systemów, procesów czy zjawisk, w szczególności w badaniu zjawisk społeczno-ekonomicznych. Właśnie dlatego w badaniu tychże zjawisk coraz częściej wykorzystuje się miary zdefiniowane na gruncie teorii informacji. W artykule zaprezentowano pojęcia i własności entropii zmiennej losowej dyskretnej. Wykorzystano entropię Shannona wraz z jej uogólnieniami (entropią Rényiego i Tsallisa) do

badania własności rozkładów prawdopodobieństw dyskretnych zmiennych losowych. Wskazano ponadto na liczne związki miar koncentracji dyskretnego rozkładu prawdopodobieństwa z entropią. W artykule zaproponowano miarę umożliwiającą ocenę stopnia koncentracji dyskretnego rozkładu prawdopodobieństwa będącą funkcją entropii Rényiego.

Słowa kluczowe: entropia Shannona, entropia Rényiego, entropia Havrda-Charvát-Daróczy-Tsallisa, miary koncentracji

APPLICATION OF SHANNON'S ENTROPY AND ITS GENERALIZATIONS FOR STUDIES ON THE DISCRETE RANDOM VARIABLE PROBABILITY DISTRIBUTION

Summary

Everywhere where there is need for identification and measurement of indeterminacy of the distributions studied we can talk about entropy. The need to measure the degree of diversity occurs in studies on numerous systems, processes and phenomena, in particular in studies on socioeconomic phenomena. That is why in studies on those phenomena the measures or models defined on grounds of the information theory are used increasingly often. The paper presents categorization of notions and characteristics of the entropy of a discrete random variable. In addition to Shannon's entropy, the Rényi's and Tsallis entropies were applied for studies on the properties of distributions in case of probabilities of the random variables. The notion of entropy stemming from thermodynamics found application in many fields of sciences. Shannon's entropy was defined on the grounds of the information theory, the Rényi's entropy is the result of generalization of the Kolmogorov-Nagumo average while the Tsallis entropy is a certain function of Rényi's entropy. The present paper unifies these approaches by presenting one general model of concentration measure that applies Rényi's entropy.

Key words: Shannon's entropy, Rényi's entropy, Tsallis entropy, concentration measure

ZBIGNIEW ŚWITALSKI

RÓWNOWAGI NA RYNKACH Z DWUSTRONNYMI PREFERENCJAMI

1. WSTĘP

W 1962 roku Gale i Shapley w pracy [5] przedstawili pewien prosty model rekrutacji kandydatów do szkół. W modelu tym poszukuje się tzw. skojarzeń stabilnych tzn. przydziałów¹ kandydatów do szkół takich, że żadna para (kandydat, szkoła) nie ma motywacji do zmiany istniejącego przydziału (zob. def. 3.1). Skojarzenia stabilne mogą być zdefiniowane w tym modelu dzięki przyjęciu założenia o dwustronnych preferencjach w układzie kandydaci-szkoła, tzn. założenia, że każdy kandydat posiada preferencje w zbiorze szkół (na przykład umie uporządkować wszystkie szkoły w kolejności od najlepszej do najgorszej), a każda szkoła posiada preferencje w zbiorze kandydatów.

Model Gale'a-Shapleya może być traktowany jako pewien model rynku, na którym kandydaci występują w roli „kupujących”, szkoły – w roli „sprzedających”, a dobra którymi się „handluje”, to miejsca w szkołach. W literaturze ekonomicznej opisano wiele innych rynków, które mają strukturę podobną do struktury wyżej opisanego rynku „szkolnego” (zob. [9]). Są to tzw. rynki z dwustronnymi preferencjami, a jako szczególnie ważny przykład może być tutaj wymieniony rynek pracy (zob. [3], [6]). Badania takich rynków nie ograniczały się jedynie do badań teoretycznych. Przeprowadzono wiele badań empirycznych dotyczących funkcjonujących w praktyce systemów kojarzenia (np. systemów rekrutacji stażystów i lekarzy do szpitali w USA i Wielkiej Brytanii, zob. [8]). Prowadzi się również intensywne badania eksperymentalne związane z zachowaniem się uczestników takich systemów – na przykład ze sposobem, w jaki określają oni swoje preferencje (zob. np. [2], [7]).

W badaniu rynków z dwustronnymi preferencjami, opisywanych za pomocą modelu Gale'a-Shapleya, wykorzystuje się głównie pojęcie skojarzenia stabilnego, które może być traktowane jako pewien rodzaj stanu równowagi na takim rynku (zob. komentarze po def. 3.1). Ze względu na brak wyraźnie określonych cen dóbr oferowanych na takim rynku, nie korzysta się natomiast tutaj z klasycznego pojęcia równowagi rynkowej. Zauważmy jednak, że (na przykład) na rynku „szkolnym” rolę cen mogłyby odgrywać minimalne liczby punktów niezbędne do dostania się do danej szkoły. Przypuśćmy, że szkoła wyższa określa minimalną liczbę punktów z kilku przedmiotów maturalnych, gwarantującą dostanie się na dany kierunek. Wówczas kandydat, który posiada odpowiednią liczbę punktów, składając podanie do tej szkoły, jak gdyby „płaci” tą liczbą

¹ Gale i Shapley w [5] użyli terminu *assignment*, który tradycyjnie w polskiej literaturze optymalizacyjnej tłumaczy się jako *przydział*.

punktów za miejsce na tym kierunku, a szkoła, przyjmując go, „sprzedaje” mu to miejsce.

W rzeczywistości systemy rekrutacji funkcjonują inaczej, gdyż dopiero po przyjęciu kandydatów możemy określić tą minimalną liczbę punktów (jako liczbę punktów, którą posiadał najgorszy z przyjętych kandydatów, przy założeniu, że wszystkie miejsca zostały „obsadzone”). Podana wyżej interpretacja pokazuje jednak, że możliwe jest określenie pewnej liczby, która odpowiada cenie dobra na tradycyjnym rynku. Możliwe jest, w związku z tym, określenie równowagi rynkowej dla rynku „szkolnego” i innych podobnych rynków (zob. def. 4.3), a więc można zastanawiać się nad związkiem tak określonej równowagi z pojęciem skojarzenia stabilnego.

Warto zwrócić uwagę, że w ostatnich latach pojawiło się wiele prac (np. [1], [10]), których autorzy badają związki między skojarzeniami stabilnymi, a równowagami rynkowymi w tzw. modelach Shapleya-Shubika (jest to pewien dyskretny model kojarzenia różny od modelu Gale’a-Shapleya). W modelach tych (w przeciwieństwie do modeli typu Gale’a-Shapleya) przyjmuje się, że preferencje kupujących zależą od cen towarów oferowanych przez sprzedających. Nie ma natomiast opracowań, w których zależności między skojarzeniami stabilnymi, a równowagami rynkowymi byłyby badane dla modeli typu Gale’a-Shapleya.

Celem artykułu jest wprowadzenie precyzyjnej, formalnej definicji stanu równowagi rynkowej w modelu Gale’a-Shapleya (ogólnie: na rynku z dwustronnymi preferencjami) oraz podanie związków między tak określoną równowagą rynkową, a skojarzeniem stabilnym. Równowaga rynkowa zostanie zdefiniowana w sposób możliwie ogólny, za pomocą pewnego układu warunków, których spełnienie gwarantuje kupującym otrzymanie towarów oferowanych przez sprzedających (zob. rozdz. 4). Takie ogólne podejście umożliwia wprowadzenie pojęcia równowagi nawet w sytuacji, gdy na danym rynku nie da się w naturalny sposób wprowadzić wielkości, które mogłyby spełniać rolę cen.

Zawartość poszczególnych rozdziałów, główne tezy oraz źródła pracy są następujące. W rozdziale 2 został przedstawiony klasyczny model rynku wzorowany na modelu Gale’a-Shapleya ([5], [9]) oraz zostały opisane dwa uproszczone modele rynków służące jako ilustracja przedstawionej teorii (model w przykładzie 2.1 jest znanym modelem Gale’a-Shapleya [5], model w przykładzie 2.2 jest oryginalny). W rozdziale 3 została wprowadzona klasyczna definicja stabilności [5]. W rozdziale 4 zostało wprowadzone ogólne pojęcie równowagi w modelu G-S zaczerpnięte z pracy [12]. W rozdziale 5 zostały zdefiniowane tzw. rozszerzenia poprawiające oraz sformułowano twierdzenie (zaczerpnięte z [12]) o wzajemnie jednoznacznej odpowiedniości między skojarzeniami stabilnymi, a tzw. równowagami stabilnymi. Rozdziały 6 i 7 zawierają niepublikowaną dotąd teorię tzw. równowag porządkowych, a w szczególności twierdzenie 7.1 o wzajemnie jednoznacznej odpowiedniości między skojarzeniami stabilnymi, a tzw. równowagami brzegowymi. Rozdział 8 zawiera przykłady ilustrujące omawianą teorię.

2. MODEL RYNKU Z DWUSTRONNYMI PREFERENCJAMI

Będziemy posługiwali się bardzo uproszczonym modelem rynku, w którym występują skończone zbiory kupujących i sprzedających, skończona liczba niepodzielnych

towarów, a także preferencje kupujących i sprzedających (kupujący mają preferencje w zbiorze sprzedających, a sprzedający – preferencje w zbiorze kupujących). W przykładach 2.1 i 2.2 nieco konkretyzujemy ten model, chociaż, ze względu na cel pracy, nie będziemy zajmowali się szczegółowym opisem faktycznie funkcjonujących rynków (ani rynku „szkolnego”, ani rynku nieruchomości). Opisy i modele różnych rynków tego rodzaju znajdują się np. w [8] i [9].

Założmy, że dany jest skończony zbiór kupujących, oznaczony przez K i skończony zbiór sprzedających, oznaczony przez S . Każdy sprzedający $s \in S$ posiada q_s ($q_s \geq 1$) niepodzielnych jednostek pewnego towaru, które może sprzedać kupującemu (różni sprzedający posiadają różne towary, natomiast każdy konkretny sprzedający posiada q_s identycznych egzemplarzy dokładnie jednego towaru). Każdy kupujący chce kupić dokładnie jedną jednostkę któregoś z towarów, a każdy sprzedający chciałby sprzedać wszystkie q_s jednostek swojego towaru (może się oczywiście zdarzyć, że dany kupujący ostatecznie nic nie kupi, a dany sprzedający nie wszystko sprzeda).

Zakładamy też, że każdy kupujący $k \in K$ ma preferencje w zbiorze S określone za pomocą pewnego liniowego porządku $>_k$. Innymi słowy, kupujący k jest w stanie ustawić wszystkich sprzedających (czyli tym samym towary, które oni posiadają) w kolejności od najbardziej mu odpowiadającego do najmniej mu odpowiadającego. Ciąg sprzedających ustawionych w takiej kolejności będzie nazywany listą preferencji kupującego k . Jeśli na przykład $S = \{s, t, u\}$, to zapis $k: tsu$ oznacza, że kupującemu k najbardziej odpowiada towar, którym dysponuje sprzedający t , następnie towar sprzedającego s , a najmniej mu odpowiada towar sprzedającego u (inaczej $t >_k s >_k u$). Ciąg tsu jest w tym przypadku listą preferencji kupującego k . Przyjmujemy, dla uproszczenia, że równoważności są niedopuszczalne, to znaczy dla dowolnych dwóch sprzedających (towarów) s_1 i s_2 , albo k woli s_1 od s_2 , albo woli s_2 od s_1 i inne sytuacje są niemożliwe. Symbol $s_1 \geq_k s_2$ będzie oznaczał, że $s_1 >_k s_2$ lub $s_1 = s_2$.

Analogicznie określamy preferencje sprzedających w zbiorze K . Zakładamy, że każdy sprzedający jest w stanie ustawić kupujących w kolejności od najbardziej mu odpowiadającego do najmniej mu odpowiadającego i utworzyć w ten sposób swoją listę preferencji zawierającą wszystkich kupujących (nie dopuszczamy równoważności). Będziemy używali symboli $>_s$ i $\geq_s$, które mają podobne znaczenie jak poprzednio (tzn. oznaczają preferencje sprzedającego $s \in S$). Oczywiście założenie o preferencjach sprzedających w stosunku do kupujących jest mniej naturalne niż założenie o preferencjach kupujących w stosunku do sprzedających, ale w podanych niżej przykładach opisane są sytuacje, w których przyjęcie tego założenia jest jak najbardziej uzasadnione.

PRZYKŁAD 2.1. (rynek „szkolny”)

Modelem rynku jest w tym przykładzie model rekrutacji kandydatów do szkół zaproponowany w 1962 roku przez Gale’a i Shapleya [5]. Model Gale’a-Shapleya stanowił inspirację do prowadzenia wielu badań (zarówno teoretycznych, jak i empirycznych) dotyczących różnego rodzaju rynków pracy czy rynków aukcyjnych ([8], [9]). W badaniach tych uwzględniano różne szczegółowe aspekty funkcjonowania takich rynków. Przedstawiony poniżej model jest tylko najprostszą wersją modelu Gale’a-Shapleya,

będącą uproszczonym opisem dobrze znanych i działających w praktyce (również w Polsce) systemów rekrutacji do szkół (np. wyższych), zob. [11], [13].

W modelu tym zbiorem K jest zbiór kandydatów do szkół. Każdy z kandydatów chciałby otrzymać („kupić”) jedno miejsce w którejś ze szkół ze zbioru S . Kandydaci mają preferencje w zbiorze S (tzn. każdy kandydat ma szkoły bardziej lub mniej mu odpowiadające i jest w stanie umieścić na swojej liście preferencji wszystkie szkoły w kolejności od najlepszej do najgorszej)². Z kolei szkoły mają preferencje w zbiorze K (szkoły punktuja kandydatów za różne osiągnięcia – na przykład uwzględniając oceny maturalne – i w ten sposób są w stanie ustawić wszystkich kandydatów na swojej liście preferencji). Zakładamy przy tym, że wszyscy kandydaci oceniani przez daną szkołę mają różne liczby punktów³. Liczbę q_s interpretujemy w tym przypadku jako limit szkoły $s \in S$, tzn. maksymalną liczbę kandydatów, którą ta szkoła jest w stanie przyjąć.

PRZYKŁAD 2.2. (rynek nieruchomości)

Ważnym nurtem badań w teorii ekonomii jest analiza rynków dóbr niepodzielnych, często o dużej wartości (domy, mieszkania, samochody, dzieła sztuki itp.). Modele matematyczne takich rynków (w szczególności modele równowagi) są często, wbrew pozorom, bardziej skomplikowane od klasycznych modeli równowagi, w których preferencje modelowane są za pomocą ciągłych funkcji użyteczności (zob. np. [14]). W naszych rozważaniach wystarczy posługiwanie się jedynie bardzo prostym modelem, w którym preferencje sprzedających są wyznaczone za pomocą cen granicznych, które kupujący gotowi są zapłacić za towary oferowane przez sprzedających.

Przypuśćmy na przykład, że sprzedającymi są deweloperzy ze zbioru S , z których każdy ($s \in S$) oferuje do sprzedaży pewną liczbę (q_s) jednakowych mieszkań (jeśli mieszkania oferowane przez s różniłyby się swoimi charakterystykami i miałyby – w oczach kupujących różną wartość, to moglibyśmy założyć, że każde mieszkanie oferuje inny sprzedawca). Kupującymi są klienci ze zbioru K , z których każdy zamierza kupić dokładnie jedno mieszkanie. Zakładamy, że każdy kupujący posiada preferencje w zbiorze oferowanych mieszkań (równoważnie – w zbiorze S) określone za pomocą pewnego liniowego porządku. Preferencje te (podobnie jak w neoklasycznej teorii konsumenta) nie zależą od cen oferowanych mieszkań. Preferencje sprzedających w zbiorze kupujących określamy następująco. Niech a_{ks} oznacza maksymalną cenę, którą kupujący $k \in K$ jest w stanie zapłacić za mieszkanie sprzedawane przez sprzedającego $s \in S$. Zakładamy, że jeśli $k \neq l$, to $a_{ks} \neq a_{ls}$. Relacja preferencji $>_s$ dla sprzedającego s jest określona w zbiorze K następująco:

² W praktyce kandydaci mają preferencje w podzbiorach zbioru S (tzn. są w stanie uporządkować tylko niektóre szkoły ze zbioru S). Przedstawioną teorię można łatwo rozszerzyć również na taki przypadek (zob. [12]).

³ W praktyce kandydaci mogą mieć tę samą liczbę punktów. Zakładamy, że w takiej sytuacji szkoła stosuje dodatkowe kryteria (np. uwzględnia punkty z dodatkowych przedmiotów), które pozwalają rozróżnić kandydatów. Jeśli szkoła nie jest w stanie rozróżnić kandydatów na swojej liście preferencji (tzn. niektórzy kandydaci są równoważni), to niezbędne byłoby wprowadzenie do teorii tzw. miękkich limitów (zob. [11]), ale nie będziemy się tym zajmowali w tym artykule.

$$k >_s l \Leftrightarrow a_{ks} > a_{ls}$$

(sprzedający woli klienta, który gotowy jest zapłacić więcej od tego, który gotowy jest zapłacić mniej).

3. SKOJARZENIA I SKOJARZENIA STABILNE

Symbol ks będzie oznaczał parę nieuporządkowaną złożoną z elementów $k \in K$ i $s \in S$, czyli zbiór dwuelementowy $\{k, s\}$. Niech $E(K, S)$ oznacza zbiór wszystkich par nieuporządkowanych złożonych z elementów zbiorów K i S , tzn.

$$E(K, S) = \{ks: k \in K, s \in S\}.$$

Zbiór wszystkich transakcji na rynku można przedstawić jako pewien podzbiór zbioru $E(K, S)$. Podzbiór taki będę nazywał skojarzeniem⁴. Jeśli $\mu \subset E(K, S)$ jest skojarzeniem, to symbol $\mu(k)$ będzie oznaczał sprzedającego, z którym k zawarł transakcję (oczywiście może być co najwyżej jeden taki sprzedający, jeśli takiego sprzedającego nie ma to przyjmujemy $\mu(k) = \emptyset$). Symbol $\mu(s)$ będzie natomiast oznaczał zbiór kupujących, którzy zawarli transakcję z s (może być $\mu(s) = \emptyset$). Z podanych wcześniej założeń wynika, że liczba elementów zbioru $\mu(s)$ nie przekracza q_s . Skojarzenie μ będzie się nazywało skojarzeniem zupełnym, jeśli liczba elementów każdego zbioru $\mu(s)$ jest równa q_s . Wszystkie skojarzenia w dalszej części artykułu będą skojarzeniami zupełnymi. Warunkiem istnienia w zbiorze $E(K, S)$ przynajmniej jednego skojarzenia zupełnego jest nierówność $n \geq \sum_{(s \in S)} q_s$, gdzie n jest liczbą elementów zbioru K .

Definicja 3.1. Skojarzenie $\mu \subset E(K, S)$ nazywa się skojarzeniem stabilnym, jeśli nie istnieje para ks (nazywana parą blokującą dla μ) taka, że

- 1) k woli s od $\mu(k)$,
- 2) s woli k od przynajmniej jednego $l \in \mu(s)$

(zakładamy, że warunek 1) jest automatycznie spełniony, jeśli $\mu(k) = \emptyset$).

Zbiór transakcji μ jest więc niestabilny, jeśli istnieje para ks (blokująca dla μ) taka, że k i s woleliby zawrzeć transakcję ze sobą niż ze swoimi aktualnymi partnerami (tzn. s z jednym ze swoich aktualnych partnerów). W takiej sytuacji k i s będą mieli tendencję do zmiany istniejącego układu (np. zerwania dotychczasowych umów). Jeśli więc chcemy, aby rynek był stabilny, to powinniśmy dążyć (lub zachęcać uczestników rynku) do tworzenia skojarzeń stabilnych.

Pojęcie skojarzenia stabilnego jest podstawą teorii kojarzenia zapoczątkowaną przez Gale'a i Shapleya ([5]) i rozwijaną przez wielu ekonomistów (zob. np. [9]). Skojarzenie stabilne może być traktowane jako pewien rodzaj stanu równowagi na rynku z dwustronnymi preferencjami (gdyż żadna para uczestników rynku nie ma

⁴ Używam terminu *skojarzenie*, który odpowiada używanemu powszechnie w teorii rynków z dwustronnymi preferencjami terminowi *matching*. W teoriach równowagi ekonomicznej często używany jest w tym kontekście termin *alokacja*.

tendencji do zmiany istniejącego układu). Zauważmy, że gdybyśmy wprowadzili na takim rynku jakiś układ cen, to moglibyśmy zdefiniować zwykłą równowagę rynkową (w sensie Walrasa). Interesujące byłoby wówczas następujące pytanie: Jaki jest związek skojarzeń stabilnych z równowagami rynkowymi wyznaczonymi przez pewien układ cen? W następnych punktach będę próbował odpowiedzieć na to pytanie, przy czym równowagę rynkową zdefiniuję w sposób bardzo ogólny, niekoniecznie odwołujący się do określonego układu cen (równowaga cenowa będzie szczególnym przypadkiem tak ogólnie określonej równowagi).

4. RÓWNOWAGA NA RYNKU Z DWUSTRONNYMI PREFERENCJAMI

Założmy, że sprzedający określają pewne warunki, których spełnienie przez kupujących gwarantuje im (kupującym) otrzymanie dóbr posiadanych przez sprzedających. Typowym warunkiem tego rodzaju jest warunek cenowy. Jeśli mianowicie sprzedający s określa cenę p_s na swoje dobro, to oznacza to, że stawia warunek: „Jeśli zapłacisz mi kwotę p_s , to otrzymasz jedną jednostkę posiadanego przeze mnie dobra”. Szkoły mogą określić pewne progi punktowe jako warunki dostania się do danej szkoły. Warunek taki można sformułować następująco: „Jeśli wykażesz się odpowiednią liczbą punktów z wyznaczonych przedmiotów maturalnych, to zostaniesz przyjęty do naszej szkoły”. Szkoły, firmy lub inne instytucje (ogólnie: dysponenti pewnych dóbr) mogą stawiać bardzo różne warunki, które gwarantują „kupującym” otrzymanie danych dóbr.

Założmy więc, że w rozważanym przez nas modelu każdy sprzedający $s \in S$ wyznaczył pewien warunek (warunki) taki, że jeśli kupujący $k \in K$ spełnił ten warunek, to może otrzymać jedną jednostkę dobra posiadanego przez s . Symbolem $W(s)$ oznaczmy zbiór kupujących, którzy spełnili warunek określony przez sprzedającego s , tzn.

$$W(s) = \{k \in K : k \text{ spełnił warunek wyznaczony przez } s\}.$$

W dalszych rozważaniach (w szczególności przy definiowaniu równowagi) istotne będą jedynie zbiory $W(s)$, a nie konkretny opis warunków wyznaczających te zbiory. Warunki wyznaczone przez sprzedających możemy więc (ryzykując pewną niedokładność, niemającą jednak wpływu na postać prezentowanych dalej definicji i twierdzeń) utożsamić ze zbiorami $W(s)$. Rodzinę zbiorów $W(s) \subset K$ indeksowaną przez elementy $s \in S$ będziemy nazywali układem warunków na rynku z dwustronnymi preferencjami (rynek DP). Wygodnie będzie przedstawiać taką rodzinę za pomocą funkcji $W: S \rightarrow P(K)$ ($P(K)$ jest rodziną wszystkich podzbiorów zbioru K). Przyjmujemy więc następującą definicję.

Definicja 4.1. Układem warunków na rynku DP nazywamy dowolną funkcję $W: S \rightarrow P(K)$.

To, że dany jest układ warunków W , oznacza że określone są pewne podzbiory $W(s)$ zbioru wszystkich kupujących K . Każdy ze zbiorów $W(s)$ wyznacza ograniczenie, jakie stawia sprzedający s kupującym, tzn. tylko kupujący ze zbioru $W(s)$ mogą

zawrzeć transakcję ze sprzedającym s . O zbiorach $W(s)$ w tej chwili nic nie zakładamy, tzn. przyjmujemy, że każdy sprzedający może w zupełnie dowolny sposób określić warunki ograniczające zbiór kupujących, z którymi gotowy jest zawrzeć transakcję (w szczególności zakładamy, że możliwy jest przypadek, gdy $W(s) = \emptyset$, tzn. warunki mogą być tak silne, że żaden kupujący nie jest w stanie ich spełnić).

W dalszej części artykułu będziemy badać, jakie rodzaje warunków (tzn. jakie rodziny zbiorów $W(s)$) wyznaczają równowagę na rynku DP , oraz co trzeba założyć o zbiorach $W(s)$, żeby równowadze tej odpowiadało skojarzenie stabilne.

Aby zdefiniować równowagę na rynku DP , musimy określić, przy danym układzie warunków W , popyt na dobra oferowane przez sprzedających. Popyt na dobro oferowane przez sprzedającego s tworzą mianowicie ci kupujący, którzy spełniają warunki $W(s)$ i dla których s jest najlepszym sprzedawcą spośród tych, których warunki dany kupujący spełnił.

Określimy wielkość popytu w sposób bardziej formalny. Niech

$$D_W(k) = \{s \in S : k \in W(s)\}$$

będzie zbiorem dóbr (sprzedających) „dopuszczalnych” dla kupującego k przy warunkach W , tzn. zbiorem sprzedających, których warunki k spełnił i którzy gotowi są w związku z tym sprzedać mu posiadane przez siebie dobro (nie wykluczamy sytuacji, że zbiór $D_W(k)$ jest pusty, tzn. że k nie spełnił warunków żadnego sprzedającego).

Dla danego kupującego k , dla którego $D_W(k) \neq \emptyset$, określamy następnie najlepszego sprzedającego jako

$$M_W(k) = \max \{s \in S : s \in D_W(k)\}.$$

W podanym wzorze bierzemy maksimum ze względu na relację preferencji kupującego k w zbiorze $D_W(k)$ (ponieważ relacja preferencji jest liniowym porządkiem w S , a więc również w $D_W(k)$, więc istnieje dla niej jednoznacznie określony element maksymalny w zbiorze $D_W(k)$).

Definicja 4.2. Mówimy, że kupujący k tworzy popyt na dobro oferowane przez sprzedającego s , jeśli $M_W(k) = s$.

Innymi słowy, k tworzy popyt na dobro oferowane przez s , jeśli s jest najlepszym sprzedającym wśród wszystkich sprzedających dopuszczalnych dla k .

Zbiór wszystkich kupujących, którzy tworzą popyt na dobro oferowane przez s oznaczymy symbolem $\varphi_W(s)$. Mamy więc

$$\varphi_W(s) = \{k \in K : M_W(k) = s\}.$$

Wielkość popytu na dobro s możemy określić jako liczbę elementów zbioru $\varphi_W(s)$ (będziemy ją oznaczać symbolem $|\varphi_W(s)|$). Ponieważ podaż dobra s jest równa q_s , dochodzimy w oczywisty sposób do pojęcia równowagi.

Definicja 4.3. Układ warunków W tworzy równowagę na rynku z dwustronnymi preferencjami, jeśli dla dowolnego $s \in S$ mamy

$$|\varphi_W(s)| = q_s.$$

W sytuacji równowagi liczba kupujących, którzy tworzą popyt na dobro oferowane przez s jest więc dokładnie równa podaży dobra s (tzn. liczbie q_s).

Zauważmy, że wprowadzone pojęcie równowagi jest analogiczne do pojęcia równowagi konkurencyjnej w klasycznych (ciągłych) modelach rynku. Zbiór $D_W(k)$ może być traktowany jako odpowiednik zbioru wiązek dopuszczalnych w teorii konsumenta, natomiast element $M_W(k)$ – jako odpowiednik pojęcia wiązki optymalnej. Funkcja $(W, s) \rightarrow |\varphi_W(s)|$ może być interpretowana jako funkcja popytu w rozważanym przez nas modelu.

5. SKOJARZENIA STABILNE A RÓWNOWAGI

Okazuje się, że istnieje ścisły związek między skojarzeniami stabilnymi, a równowagami rynkowymi zdefiniowanymi za pomocą definicji 4.3. Aby określić ten związek, wprowadzimy dwa pojęcia: równowagi podstawowej i rozszerzenia poprawiającego daną równowagę. Zauważmy najpierw dwa proste fakty dotyczące zbiorów $\varphi_W(s)$:

- (F1) Zbiory $\varphi_W(s)$ są rozłączne (dla różnych s , przy ustalonym W),
- (F2) $\varphi_W(s) \subset W(s)$ dla dowolnego $s \in S$.

Fakt (F1) wynika stąd, że dla każdego k optymalny sprzedający $M_W(k)$ jest jednoznacznie określony, a fakt (F2) jest konsekwencją definicji $\varphi_W(s)$, $M_W(k)$ i $D_W(k)$.

Definicja 5.1. Równowaga W jest równowagą podstawową jeśli dla dowolnego $s \in S$ mamy

$$\varphi_W(s) = W(s).$$

Jeśli warunki W tworzą równowagę podstawową, to oznacza to, że:

1. Każdy kupujący, który spełnia warunki $W(s)$, tworzy popyt na dobro s (nie ma kupujących, którzy spełnialiby $W(s)$ i chcieliby otrzymać inne dobro niż s),
2. Każdy kupujący spełnia warunki tylko jednego sprzedającego (bo zbiory $W(s)$ są rozłączne, co wynika z (F1) i definicji 5.1)

Można też powiedzieć, że w warunkach równowagi podstawowej każdy sprzedający sprzedaje swój towar wybranym kupującym, którzy są z góry ustaleniu (sprzedający s sprzedaje swój towar wszystkim ze zbioru $W(s)$).

Równowagi podstawowe są nieco sztucznym tworem (na przykład na rynku szkolnym na ogół będzie wielu kandydatów spełniających warunki danej szkoły, ale nie pragnących uczyć się w tej szkole), będą jednak potrzebne w dalszych rozważaniach, gdyż każdej takiej równowadze odpowiada dokładnie jedno skojarzenie i na odwrót.

Niech bowiem $\mu \subset E(K, S)$ będzie dowolnym skojarzeniem. Możemy wówczas określić równowagę podstawową wzorem:

$$W(\mu)(s) = \mu(s) = \{k \in K : ks \in \mu\}.$$

Okazuje się, że odwzorowanie $\mu \rightarrow W(\mu)$ jest wzajemnie jednoznaczna odpowiedniością między zbiorem skojarzeń (zupełnych), a zbiorem równowag podstawowych (zob. [12]).

Interesująca byłaby odpowiedź na pytanie: jakie równowagi odpowiadają przy odwzorowaniu $\mu \rightarrow W(\mu)$ skojarzeniom stabilnym. Odpowiedź na to pytanie jest zawarta w poniżej podanym twierdzeniu 5.4. Przed sformułowaniem tego twierdzenia potrzebne są jeszcze dwie definicje.

Definicja 5.2. Niech W będzie równowagą podstawową. Rozszerzeniem poprawiającym równowagę W nazywamy odwzorowanie $V: S \rightarrow P(K)$ takie, że

1. $W(s) \subset V(s)$ dla dowolnego $s \in S$,
2. Dla dowolnego $s \in S$, jeśli k spełnia warunki $V(s)$ i nie spełnia warunków $W(s)$, to istnieje $l \in W(s)$ taki, że k jest lepszy od l dla sprzedającego s .

Określenie rozszerzenia poprawiającego równowagę W oznacza, dla każdego sprzedającego, osłabienie warunków W , ale w taki sposób, że każdy kupujący spełniający nowe warunki i niespełniający starych jest lepszy od przynajmniej jednego kupującego spełniającego stare warunki. Inaczej mówiąc, oznacza to, że każdy sprzedający określa nowe, mniej restrykcyjne warunki zakupu swoich towarów, przy czym stara się to zrobić w taki sposób, aby każdy kupujący spełniający te nowe warunki był dla tego sprzedającego co najmniej tak dobry, jak najgorszy z dotychczasowych kupujących (na przykład, żeby oferował cenę równą co najmniej najniższej z cen dotychczas oferowanych). Rozszerzenie warunków nie powinno więc być takie, aby którykolwiek ze sprzedających na tym stracił.

Nie każde rozszerzenie poprawiające daną równowagę jest równowagą (np. rozszerzenie $g(v)$ w przykładzie 8.1). Równowagi, które mają tę własność, że każde ich rozszerzenie poprawiające jest równowagą, będę dalej nazywał równowagami stabilnymi. Mamy więc następującą definicję.

Definicja 5.3. Równowaga podstawowa nazywa się równowagą stabilną, jeśli dowolne jej rozszerzenie poprawiające jest równowagą.

W tradycyjnym ujęciu równowagi (np. w fizyce lub w teorii równowagi rynkowej) stabilność równowagi oznacza w pewnym sensie trudność w wytrąceniu danego układu ze stanu równowagi. W rozważanym tutaj modelu dyskretnym oczywiście trudno byłoby wprowadzić formalnie analogiczne pojęcie stabilności. Wydaje się jednak, że powyższa definicja jest uprawniona i ma pewien element wspólny z klasyczną definicją, gdyż mówi o tym, że jeśli przejdziemy od danej równowagi do jej rozszerzenia poprawiającego (czyli dokonamy operacji, którą w pewnym sensie można nazwać „wytrąceniem ze stanu równowagi”), to znowu znajdziemy się w stanie równowagi.

Okazuje się, że istnieje pewien prosty związek między tak zdefiniowanymi równowagami stabilnymi, a skojarzeniami stabilnymi w sensie Gale'a-Shapleya (zob. [12]).

Twierdzenie 5.4. Odwzorowanie $\mu \rightarrow W(\mu)$ określa wzajemnie jednoznaczną odpowiedniość między zbiorem skojarzeń stabilnych, a zbiorem równowag stabilnych.

Z twierdzenia 5.4. wynika, że każdemu skojarzeniu stabilnemu odpowiada dokładnie jedna równowaga stabilna, a każdej równowadze stabilnej odpowiada dokładnie jedno skojarzenie stabilne. Z twierdzenia tego możemy też otrzymać następującą charakteryzację skojarzeń stabilnych:

Wniosek 5.5. Skojarzenie μ jest stabilne wtedy, i tylko wtedy, gdy równowaga $W(\mu)$ jest stabilna.

6. RÓWNOWAGI PORZĄDKOWE

Zdefiniujemy teraz tzw. równowagi porządkowe, których szczególnym rodzajem są równowagi cenowe (tzn. równowagi, dla których warunki $W(s)$ określone są za pomocą cen p_s wyznaczonych przez sprzedających, a preferencje sprzedających w zbiorze K określone są za pomocą maksymalnych cen a_{ks} oferowanych przez kupujących – podobnie jak w przykładzie 2.2).

Równowaga porządkowa jest określona za pomocą tzw. warunków porządkowych, tzn. warunków, które mają tę własność, że jeśli jakiś kupujący spełnia taki warunek, to każdy kupujący lepszy od niego też spełnia ten warunek. Jeśli na przykład sprzedający s sprzedaje swój towar za cenę p_s (co oznacza, że zapłacenie ceny p_s jest warunkiem otrzymania przez kupującego towaru oferowanego przez s), a pewien kupujący $k \in K$ jest skłonny zapłacić cenę $a_{ks} \geq p_s$, to każdy kupujący $l \in K$, który gotowy jest zapłacić więcej niż a_{ks} (np. $a_{ls} > a_{ks}$) automatycznie spełnia warunek sprzedającego s , bo oczywiście $a_{ls} > p_s$.

Definicja 6.1. Warunek $W(s)$ nazywa się warunkiem porządkowym, jeśli dla dowolnych $k, l \in K$ zachodzi:

$$k \in W(s) \text{ i } l >_s k \Rightarrow l \in W(s).$$

Jeśli $W(s) \neq \emptyset$, to „porządkowość” $W(s)$ oznacza, że

$$W(s) = \{k \in K : k \geq_s \min W(s)\}.$$

(Jeśli $W(s) \neq \emptyset$, to $\min W(s)$ oznacza minimalny element zbioru $W(s)$ ze względu na porządek $>_s$). Innymi słowy, $W(s)$ jest w tym przypadku przedziałem początkowym w zbiorze liniowo uporządkowanym K . Przedział ten składa się z pewnego „granicznego” kupującego $k = \min W(s)$ i wszystkich kupujących lepszych od niego (w sensie relacji $>_s$). Graniczny (tzn. najgorszy w zbiorze $W(s)$) kupujący będzie nazywany

minimalnym kupującym dla s ze względu na warunki $W(s)$ (minimalny kupujący jest określony tylko w przypadku, gdy $W(s) \neq \emptyset$).

Przykładem warunku porządkowego jest właśnie przedstawiony powyżej warunek cenowy, dla którego

$$W(s) = \{k \in K : a_{ks} \geq p_s\},$$

gdzie p_s jest ceną określoną przez sprzedającego $s \in S$, a liczby a_{ks} są interpretowane tak jak w przykładzie 2.2.

Definicja 6.2. Równowaga W nazywa się równowagą porządkową, jeśli wszystkie warunki $W(s)$, $s \in S$, są warunkami porządkowymi.

Definicja 6.3. Równowaga porządkowa W nazywa się równowagą brzegową, jeśli każdy minimalny kupujący dla $s \in S$, przy warunkach $W(s)$, tworzy popyt na s , tzn. jeśli dla dowolnego $k \in K$ spełniony jest warunek:

$$k = \min W(s) \Rightarrow s = M_W(k).$$

Innymi słowy, dla równowagi brzegowej zachodzi $\min W(s) = \min \varphi_W(s)$ (zauważmy, że $W(s) \neq \emptyset$, bo W jest równowagą). Warunki $W(s)$ dla równowagi brzegowej mają więc tę własność, że jeśli pewien kupujący k jest najgorszy w zbiorze wszystkich kupujących spełniających warunki $W(s)$, to sprzedający s jest najlepszy dla k w zbiorze wszystkich sprzedających, których warunki kupujący k spełnił.

7. ZWIĄZEK RÓWNOWAG PORZĄDKOWYCH ZE SKOJARZENIAMI STABILNYMI

Twierdzenie 5.4 pokazuje, że istnieje ścisły związek między skojarzeniami stabilnymi a równowagami stabilnymi (będącymi pewnym rodzajem równowag podstawowych). Okazuje się, że można ustanowić również podobny związek (podany niżej w twierdzeniu 7.1) między skojarzeniami stabilnymi a równowagami brzegowymi (będącymi pewnym rodzajem równowag porządkowych).

Zdefiniujemy najpierw pewne odwzorowanie ze zbioru skojarzeń (zupełnych) do zbioru funkcji $S \rightarrow P(K)$. Niech μ będzie dowolnym skojarzeniem. Definiujemy funkcję $g(\mu) : S \rightarrow P(K)$ następująco:

$$g(\mu)(s) = \{k \in K : k \geq_s \min \mu(s)\}.$$

Ponieważ μ jest zupełne, więc zbiory $\mu(s)$ są niepuste, czyli odwzorowanie g jest dobrze określone.

Twierdzenie 7.1. Odwzorowanie $\mu \rightarrow g(\mu)$ określa wzajemnie jednoznaczną odpowiedniość między zbiorem skojarzeń stabilnych a zbiorem równowag brzegowych.

Dowód. Określimy odwzorowanie h ze zbioru funkcji $S \rightarrow P(K)$ do zbioru podzbiorów zbioru $E(K, S)$. Niech V będzie funkcją $S \rightarrow P(K)$. Definiujemy zbiór $h(V) \subset E(K, S)$ następująco:

$$h(V) = \{ks \in E(K, S) : s = M_V(k)\} = \{ks \in E(K, S) : k \in \varphi_V(s)\}.$$

Łatwo zauważyć, że jeśli V jest równowagą, to $h(V)$ jest skojarzeniem (zupełnym). Aby udowodnić twierdzenie 7.1 wystarczy pokazać, że odwzorowanie g przekształca skojarzenia stabilne na równowagi brzegowe, h przekształca równowagi brzegowe na skojarzenia stabilne oraz, że g i h są wzajemnie do siebie odwrotne. W tym celu udowodnimy następujące dwa lematy:

Lemat 7.2. Jeśli μ jest skojarzeniem stabilnym, to $g(\mu)$ jest równowagą brzegową oraz $h(g(\mu)) = \mu$,

Lemat 7.3. Jeśli V jest równowagą brzegową, to $h(V)$ jest skojarzeniem stabilnym oraz $g(h(V)) = V$.

Dowód lematu 7.2. Niech μ będzie dowolnym skojarzeniem stabilnym. Udowodnimy po kolei następujące fakty:

- (1) $g(\mu)$ jest równowagą porządkową,
- (2) $\mu \subset h(g(\mu))$,
- (3) $\mu = h(g(\mu))$,
- (4) $g(\mu)$ jest równowagą brzegową.

Aby udowodnić (1), należy zauważyć, że $g(\mu)$ jest rozszerzeniem poprawiającym równowagi podstawowej $W(\mu)$ (bo $W(\mu)$ jest wyznaczona za pomocą zbiorów $\mu(s)$), a więc na mocy wniosku 5.5 i definicji 5.3, uwzględniając założenie, że μ jest stabilne, otrzymujemy, że $g(\mu)$ jest równowagą. Bezpośrednio z definicji $g(\mu)$ wynika też, że $g(\mu)$ jest równowagą porządkową.

Przystępujemy do dowodu zawierania (2). Załóżmy, że $ks \in \mu$. Stąd $k \in \mu(s)$, czyli $k \geq_s \min \mu(s)$, a więc $k \in g(\mu)(s)$. Stąd $s \in D_{g(\mu)}(k)$. Niech $t = M_{g(\mu)}(k)$ będzie elementem maksymalnym w zbiorze $D_{g(\mu)}(k)$ ze względu na relację $>_k$ (taki element istnieje, bo $D_{g(\mu)}(k) \neq \emptyset$). Udowodnimy, że $t = s$. Dowód przeprowadzimy nie wprost. Gdyby $t \neq s$, to $t >_k s$ (bo t jest maksymalny), a więc, z jednoznaczności $\mu(k)$, $kt \notin \mu$, czyli $k \notin \mu(t)$. Dalej, ponieważ $t \in D_{g(\mu)}(k)$, więc $k \in g(\mu)(t)$, a stąd $k \geq_t \min \mu(t)$. Ponieważ $k \notin \mu(t)$, więc $k >_t \min \mu(t)$. Stąd i z warunku $t >_k s$ wynika, że para kt spełnia warunki 1) i 2) z definicji 3.1 (dla skojarzenia μ), a więc skojarzenie μ nie jest stabilne, wbrew założeniu. Mamy więc $t = s$, czyli $s = M_{g(\mu)}(k)$, a więc $ks \in h(g(\mu))$. Udowodniliśmy w ten sposób zawieranie $\mu \subset h(g(\mu))$.

Aby udowodnić (3), czyli równość $\mu = h(g(\mu))$ wystarczy zauważyć, że zbiory μ oraz $h(g(\mu))$ są równoliczne. Wynika to stąd, że $g(\mu)$ jest równowagą oraz stąd, że dla dowolnej równowagi V , skojarzenie $h(V)$ zawiera dokładnie $\sum_s q_s$ elementów, tzn. tyle samo co skojarzenie μ (bo μ jest zupełne).

Warunek (4) wynika teraz w prosty sposób z równania (3). Mamy bowiem dla dowolnego $s \in S$:

$$\mu(s) = h(g(\mu))(s) = \varphi_{g(\mu)}(s).$$

Stąd $\min g(\mu)(s) = \min \mu(s) = \min \varphi_{g(\mu)}(s)$, a więc równowaga $g(\mu)$ jest brzegowa.

Dowód lematu 7.3. Udowodnimy najpierw, że dla dowolnej równowagi porządkowej V , zbiór $h(V)$ jest skojarzeniem stabilnym. Niech $\mu = h(V)$. Z definicji $h(V)$ wynika, że μ jest skojarzeniem. Załóżmy, że μ jest niestabilne. Istnieje więc para ks taka, że $s >_k \mu(k)$ oraz $k >_s l$ dla pewnego $l \in \mu(s)$. Ponieważ $l \in \mu(s) = h(V)(s) = \varphi_V(s) \subset V(s)$, a równowaga V jest porządkowa, więc $k >_s l$ implikuje $k \in V(s)$. Stąd $s \in D_V(k)$. Mamy więc $D_V(k) \neq \emptyset$ oraz $\mu(k) = h(V)(k) = M_V(k) \geq_k s$, a to jest sprzeczne z przyjętym na początku warunkiem $s >_k \mu(k)$. Wynika stąd, że skojarzenie $\mu = h(V)$ jest stabilne.

Udowodnimy teraz, że jeśli V jest równowagą brzegową, to $g(h(V)) = V$. Korzystając z definicji 6.1 i 6.3 oraz z definicji funkcji g i h , dla dowolnego $s \in S$ mamy:

$$\begin{aligned} g(h(V))(s) &= \{k \in K : k \geq_s \min h(V)(s)\} = \{k \in K : k \geq_s \min \varphi_V(s)\} \\ &= \{k \in K : k \geq_s \min V(s)\} = V(s), \end{aligned}$$

co kończy dowód lematu 7.3, a więc również twierdzenia 7.1.

8. PRZYKŁADY

Podamy teraz dwa przykłady ilustrujące przedstawioną teorię.

PRZYKŁAD 8.1.

Niech $K = \{A, B, C, D, E\}$ oraz $S = \{X, Y, Z\}$ będą zbiorami kupujących i sprzedających odpowiednio. Załóżmy, że limity dla sprzedających są równe: $q_X = 2$, $q_Y = 2$, $q_Z = 1$, a preferencje kupujących i sprzedających mają postać $A: XYZ$, $B: YZX$, $C: YXZ$, $D: ZYX$, $E: ZXY$, $X: CEBAD$, $Y: ACBED$, $Z: BECDA$. Rozważmy skojarzenie $\mu = \{AX, BY, CY, DX, EZ\}$. Jest to skojarzenie stabilne, ponieważ kupujący A, B, C, E nie mogą tworzyć par blokujących (są skojarzeni z najlepszymi sprzedającymi na swoich listach preferencji), skąd wynika, że jedynymi parami blokującymi mogłyby być pary DZ i DY . Zauważmy jednak, że Z woli E (z którym jest skojarzony) od D , a Y woli C i B od D . Stąd DZ i DY nie są parami blokującymi. Skojarzeniu μ odpowiada równowaga podstawowa

$$W = W(\mu) = \{AD, CB, E\}$$

(powyższy zapis oznacza, że $W(X) = \{A, D\}$, $W(Y) = \{C, B\}$, $W(Z) = \{E\}$). Mamy więc tutaj przykład odpowiedniości $\mu \rightarrow W(\mu)$, tak jak w twierdzeniu 5.4. Na mocy tego twierdzenia równowaga $W(\mu)$ jest stabilna, a więc każde jej rozszerzenie poprawiające jest równowagą.

Przykładami równowag niepodstawowych są:

$$U = \{BAD, ACB, BE\},$$

$$V = \{CEBAD, ACB, BE\},$$

$$T = \{CEBAD, ACBE, BEC\}.$$

Wszystkie te równowagi wyznaczają to samo skojarzenie μ , inaczej mówiąc, $h(U) = h(V) = h(T) = \mu$. Równowagi U i V są rozszerzeniami poprawiającymi równowagę $W(\mu)$ (a więc są równowagami na mocy twierdzenia 5.4 i definicji 5.3). Równowaga V jest brzegowa, przy czym $g(\mu) = V$. Mamy więc odpowiedniość $\mu \rightarrow g(\mu) = V$, taką jak w twierdzeniu 7.1. Równowaga T jest równowagą porządkową, chociaż nie jest brzegowa (np. kupujący E jest kupującym minimalnym dla sprzedającego Y , ale nie tworzy popytu na Y). Równowaga T jest jednocześnie przykładem rozszerzenia niepoprawiającego dla $W(\mu)$ (bo np. kupujący E jest dla Y gorszy od wszystkich kupujących, którzy spełniają warunek $W(\mu)$ wyznaczony przez Y , tzn. od C i B).

Przykładem skojarzenia niestabilnego dla podanego układu preferencji jest skojarzenie

$$v = \{AY, BY, CX, DZ, EX\}.$$

Skojarzenie v jest niestabilne, gdyż np. para CY jest parą blokującą (C woli Y od X , a jednocześnie Y woli C od B). Zauważmy, że funkcja

$$g(v) = \{CE, ACB, BED\}$$

nie jest w tym przypadku równowagą, gdyż ani C , ani E nie tworzą popytu na X (dla warunków $g(v)$). Tym bardziej więc $g(v)$ nie jest w tym przypadku równowagą brzegową.

PRZYKŁAD 8.2.

W przykładzie tym określone zostanie skojarzenie niestabilne v , dla którego $g(v)$ jest równowagą brzegową.

Załóżmy, że $K = \{A, B, C, D\}$, $S = \{X, Y\}$ są zbiorami kupujących i sprzedających, podaże sprzedających są równe $q_X = 2$, $q_Y = 2$, a preferencje określone są następująco: A : YX , B : XY , C : YX , D : XY , X : $DCBA$, Y : $CDAB$. Rozważmy skojarzenia: $\mu = \{AY, BX, CY, DX\}$, $v = \{AY, BX, CX, DY\}$. Łatwo sprawdzić, że μ jest stabilne, a v niestabilne (np. para CY jest blokująca dla v). Odpowiednie równowagi podstawowe mają postać:

$$W(\mu) = \{DB, CA\},$$

$$W(v) = \{CB, DA\}.$$

Rozszerzenia porządkowe równowag $W(\mu)$ i $W(v)$ są jednakowe i mają postać:

$$g(\mu) = g(v) = \{DCB, CDA\}.$$

Są to więc równowagi brzegowe (bo μ jest stabilne). Oczywiście $h(g(\mu)) = h(g(v)) = \mu$.

Może się więc okazać, że równowaga brzegowa jest obrazem, przy odwzorowaniu g , jakiegoś skojarzenia niestabilnego (choć, oczywiście, zawsze jest też obrazem jednoznacznie określonego skojarzenia stabilnego, co wynika z twierdzenia 7.1). Wynika stąd, że nie da się z twierdzenia 7.1 wyciągnąć wniosku analogicznego do wniosku 5.5 (wynikającego z twierdzenia 5.4), tzn. nie jest prawdą, że skojarzenie μ jest stabilne wtedy, i tylko wtedy, gdy $g(\mu)$ jest równowagą brzegową.

9. PODSUMOWANIE

W pracy zostały określone związki między skojarzeniami stabilnymi w modelu Gale'a-Shapleya a odpowiednio określonymi równowagami rynkowymi. Pojęcie równowagi zostało zdefiniowane w bardzo ogólny sposób – za pomocą układu warunków, które muszą spełniać kupujący, jeśli chcą nabyć dobra oferowane przez sprzedających. Szczególnym rodzajem tego typu warunków są warunki cenowe, występujące w klasycznych modelach równowagi rynkowej. Uzyskane wyniki mogą służyć do badania równowag w dyskretnych modelach rynków za pośrednictwem teorii skojarzeń stabilnych. Pierwszy oczywisty wniosek jest taki, że w przedstawionym modelu rynku istnieje przynajmniej jedna równowaga brzegowa (a więc również cenowa) – bo w modelu Gale'a-Shapleya zawsze istnieje co najmniej jedno skojarzenie stabilne. Jeśli z kolei, przy danym układzie preferencji, skojarzeń stabilnych jest odpowiednio dużo (w niektórych przypadkach liczba ta może zależeć nawet wykładniczo od liczby uczestników rynku – zob. [4]), to mamy gwarancję istnienia odpowiednio dużej liczby równowag brzegowych (czyli istotnie różnych, tzn. dających różne skojarzenia, równowag cenowych).

Uniwersytet Zielonogórski, Wydział Matematyki, Informatyki i Ekonometrii

LITERATURA

- [1] Camina E., [2006], *A generalized assignment game*, „Mathematical Social Sciences”, Vol. 52, s. 152-161.
- [2] Chen Y., Sönmez T., [2006], *School choice: an experimental study*, „Journal of Economic Theory”, Vol. 127, s. 202-231.
- [3] Crawford V.P., Knoer E.M., [1981], *Job matching with heterogeneous firms and workers*, „Econometrica”, Vol. 49, s. 437-450.
- [4] Gale D., [2001], *The two-sided matching problem. Origin, development and current issues*, „International Game Theory Review”, Vol. 3, s. 237-252.
- [5] Gale D., Shapley L.S., [1962], *College Admissions and the Stability of Marriage*, „American Mathematical Monthly”, Vol. 69, s. 9-15.
- [6] Kelso A.S., Crawford V.P., [1982], *Job matching, coalition formation and gross substitutes*, „Econometrica”, Vol. 50, s. 1483-1504.

- [7] Pais J., Pinter A., [2008], *School choice and information: An experimental study on matching mechanisms*, Games and Economic Behavior, Vol. 64, s. 303-328.
- [8] Roth A.E., [2002], *The economist as engineer: game theory, experimentation, and computation as tools for design economics*, „Econometrica”, Vol. 70, s. 1341-1378.
- [9] Roth A.E., Sotomayor M.A., [1992], *Two-sided matching. A study in game-theoretic modeling and analysis*, Cambridge University Press.
- [10] Sotomayor M., [2007], *Connecting the cooperative and competitive structures of the multiple-partners assignment game*, „Journal of Economic Theory”, Vol. 134, s. 155-174.
- [11] Świtalski Z., [2005], *Optymalny system rekrutacji kandydatów do szkół*, „Badania Operacyjne i Decyzje”, Vol. 3-4, s. 85-98.
- [12] Świtalski Z., [2008], *Stability and equilibria in the matching models*, „Scientific Research of the Institute of Mathematics and Computer Science”, Częstochowa University of Technology, Vol. 2(7), s. 77-85.
- [13] Świtalski Z., [2008], *O kojarzeniu małżeństw i rekrutacji kandydatów do szkół*, „Wiadomości Matematyczne”, Vol. 44, s. 35-46.
- [14] Yang Z., [2003], *A competitive market model for indivisible commodities*, „Economics Letters”, Vol. 78, s. 41-47.

Praca wpłynęła do redakcji w lutym 2010 r.

RÓWNOWAGI NA RYNKACH Z DWUSTRONNYMI PREFERENCJAMI

Streszczenie

W artykule wprowadzono pojęcie równowagi rynkowej w modelu rynku z dwustronnymi preferencjami (model jest wzorowany na znanym modelu rekrutacji kandydatów do szkół, pochodzącym od Gale'a i Shapleya z 1962 roku). Zbadano związki wprowadzonego pojęcia równowagi z pojęciem skojarzenia stabilnego zdefiniowanego przez Gale'a i Shapleya.

Słowa kluczowe: równowaga rynkowa, dwustronne preferencje, skojarzenie stabilne

EQUILIBRIA IN MARKETS WITH TWO-SIDED PREFERENCES

Summary

In the paper we have introduced a concept of market equilibrium in the model of market with two-sided preferences (the model is based on the well-known Gale-Shapley model of college admissions from 1962). We have investigated relationships between our concept of equilibrium and the concept of stable matching as defined by Gale and Shapley.

Key words: Market equilibrium, Two-sided preferences, Stable matching, College admissions, Gale-Shapley model

EWA DRABIK

WYKORZYSTANIE REGUŁ AUKCYJNYCH DO HANDLU ENERGIĄ W POLSCE

1. WSTĘP

Rynek energii w Polsce nabrał charakteru konkurencyjnego w końcu lat 90. XX wieku. Wówczas to w większości krajów europejskich uchwalono nowe prawo (w Polsce w 1997 roku), które umożliwiło tworzenie wewnętrznych rynków energii elektrycznej. Pod koniec 1999 roku powstała w Polsce Towarowa Giełda Energii Elektrycznej (TGE), która niemalże od początku swego istnienia spełniała m.in. rolę mechanizmu rynkowego służącego do wyznaczania obiektywnej ceny energii. Dzięki poparciu i kontroli Komisji Nadzoru Finansowego transakcje tam zawierane można uznać za bezpieczne.

Jako pierwszy na TGE uruchomiony został rynek „spot” (kasowy) energii elektrycznej, na którym ceny ustalane są na zasadzie reguł aukcji dwustronnej. Aukcja dwustronna jest taką formą wymiany towarów i pieniędzy, podczas której oferty składane są zarówno przez kupujących, jak i sprzedających. W każdej ofercie oprócz ceny podawana jest ilość towaru, który kontrahenci zamierzają nabyć lub sprzedać.

Badania oraz eksperymenty symulacyjne dotyczące funkcjonowania aukcji dwustronnych prowadzone są na szeroka skalę. Zapoczątkował je w latach 60. XX wieku, późniejszy laureat Nagrody Nobla, Vernon L. Smith. Próby modelowania mechanizmów aukcji dwustronnej oraz zachowań jej uczestników podejmowało wielu autorów, takich jak: Wilson (1987), Friedman (1991), Easley i Ledyard (1993), Gjestead i Dichaut (1998), Sadierih (1998) oraz Xia, Stallaert, Whinston (2004). Celem pracy jest dopasowanie jednego z istniejących modeli aukcji dwustronnej do opisu funkcjonowania TGE. Do tego celu zostanie wykorzystany model Sadieriha, który dobrze oddaje reguły aukcyjne. Zostanie również zaprezentowany algorytm wyznaczania ceny transakcji dla aukcji dwustronnych (Drabik 1999), przy czym wyniki działania tego algorytmu zostaną porównane z rzeczywistymi cenami transakcji, które zostały odnotowane na TGE.

Praca zorganizowana jest w następujący sposób. W rozdziale drugim zaprezentowane zostaną zasady funkcjonowania TGE oraz sposoby zawierania transakcji na rynku typu spot. W rozdziale trzecim omówiony zostanie jeden z modeli aukcji dwustronnej i zostanie podjęta próba jego modyfikacji, a następnie dopasowania modelu do reguł obowiązujących podczas handlu na TGE. Zostanie również zaprezentowany stochastyczny algorytm wyznaczania ceny transakcji dla aukcji dwustronnej autorstwa E. Drabik (1999). Rozdział czwarty poświęcony jest analizie danych empirycznych, pod kątem porównania cen transakcji obliczanych za pomocą algorytmu z rzeczywistymi cenami transakcji notowanymi na TGE.

2. REGUŁY FUNKCJONOWANIA TOWAROWEJ GIEŁDY ENERGII W WARSZAWIE

2.1. HISTORIA POWSTANIA I ZAKRES DZIAŁALNOŚCI TGE W WARSZAWIE

Towarowa Giełda Energii zajmuje się handlem energią elektryczną, paliwami ciekłymi i gazowymi, limitami wielkości produkcji energii elektrycznej, limitami wielkości emisji zanieczyszczeń, w szczególności CO₂, prawami majątkowymi dotyczącymi energii elektrycznej, paliw ciekłych lub gazowych. Działa ona od 7 grudnia 1999 roku, a jej powstanie jest odpowiedzią na nowe prawo międzynarodowe dotyczące handlu energią obowiązujące od końca lat 90. XX wieku w większości krajów europejskich. Jako pierwszy uruchomiony został rynek spot (kasowy) energii elektrycznej. Ceny transakcji na tym rynku ustalane są zgodnie z regułami aukcji dwustronnej, co jest głównym tematem rozważań niniejszej pracy. W 2003 roku TGE uzyskała licencję Komisji Nadzoru Finansowego na prowadzenie giełdy towarowej. W ten sposób uregulowano funkcjonowanie rynku, a także podjęto próbę wyeliminowania ewentualnych patologii spekulacyjnych. Uruchomiony został rynek praw majątkowych, w którym producenci energii odnawialnej oraz podmioty zobowiązane do zakupu świadectw pochodzenia, mogą handlować prawami majątkowymi do tych świadectw. W 2006 roku TGE uruchomiła również rynek spot dla uprawnień emisją CO₂, na którym uczestnicy mogą handlować jednostkami EUA (European Union Allowance). Dwa lata później, w 2008 roku TGE uruchomiła terminowy rynek energii elektrycznej. Kontrakty terminowe na dostawę energii elektrycznej, notowane na tym rynku, pozwalają wyznaczyć cenę energii elektrycznej w dłuższym horyzoncie czasowym, a to z kolei, umożliwia uczestnikom optymalizację kosztów sprzedaży/ zakupu energii elektrycznej. TGE spełnia także rolę mechanizmu rynkowego, który pozwala na wyznaczanie obiektywnej rynkowej ceny energii. Dzieje się tak, ponieważ reguły obowiązujące podczas zawierania transakcji są ściśle określone i przejrzyste, zaś wszyscy uczestnicy rynku, niezależnie od wielkości, są traktowani jednakowo. Wprowadzono także bardzo nowatorskie, jak na polskie warunki, rozwiązania w zakresie obrotu towarami giełdowymi. Przyszłościowo TGE postawiła sobie za cel wdrożenie nowoczesnej metody międzynarodowego handlu energią elektryczną. W systemie tym zdolności przesyłowe energii są udostępniane poprzez transakcje na współpracujących ze sobą giełdach energii. TGE dąży również do tego, aby zwiększyć interaktywność na rynku energii, a także zwiększyć swoją rolę w handlu tzw. białą energią, potocznie zwaną energią odnawialną. Obecnie TGE prowadzi następujące rynki: rynek dnia następnego (RDN), rynek gazu (RG), rynek terminowy (RT), rynek praw majątkowych, rynek uprawnień do emisji CO₂.

2.2. REGUŁY ZAWIERANIA TRANSAKCJI NA TGE; SKŁADANIE ZLECEŃ I SYSTEM NOTOWAŃ

W niniejszym punkcie zostaną przedstawione niektóre sposoby składania i realizacji zleceń na TGE, a także zaprezentowany zostanie sposób ustalania ceny energii. Rozważania dotyczą tzw. rynku dnia następnego, na którym zlecenia zarówno kupna, jak i sprzedaży składane są sekwencyjnie w określonych przedziałach czasu, które można traktować jako kolejne cykle.

Zlecenia na TGE przyjmowane są zgodnie z regulaminem na kilka sposobów. Mogą być one zbierane blokowo lub godzinowo. Przykładowy sposób składania zleceń jest następujący. W godzinach od 07:15 do 14:30, na dwa dni przed dniem dostawy, przyjmowane są zlecenia, które można, zgodnie z regulaminem, usuwać i modyfikować. Do godz. 16:00, na dwa dni przed dniem dostawy, następuje rozliczenie wartościowe transakcji zawartych na RDN. Na dzień przed dniem dostawy, w godzinach od 07:15 do 08:00 również przyjmowane są zlecenia. Po godzinie 8:00 określane są tzw. kursy RND, czyli jednolita cena transakcji za 1 MWh energii elektrycznej w danej godzinie dnia dostawy, po czym wyniki notowań podawane są na niepublicznej stronie internetowej. W godzinach 08:01 do 10:30 przyjmowane są oferty kupna/ sprzedaży w systemie notowań ciągłych. Operacje przyjmowania zleceń i rozliczania transakcji odbywa się za pośrednictwem programu komputerowego, który działa na zasadach stworzonych przez jedną z bardziej znanych firm obsługujących giełdy towarowe i towarowo-pieniężne w Ameryce i Europie.

Zlecenia składane są przez tzw. członka giełdy, czyli podmiot, który zawarł z giełdą umowę o członkostwo i został dopuszczony do „działalności giełdowej”. Typowe zlecenie powinno zawierać następujące informacje:

- a) rodzaj zlecenia (kupno/ sprzedaż),
- b) ilość energii będącej przedmiotem zlecenia,
- c) limit ceny wyrażony w PLN/ MWh lub polecenie wykonania zlecenia bez limitu ceny,
- d) termin ważności zlecenia,
- e) warunki realizacji zlecenia:
 - *dzienne* (ang. *rest of day*), które mówi, że zlecenie powinno być zrealizowane do końca bieżącego dnia;
 - *do końca okresu notowań* (ang. *good until expiry*);
 - *zlecenie do dnia* (ang. *good until date*);
 - *zlecenie czasowe* (ang. *time order*);
 - *zlecenie tylko na aukcję* (ang. *call auction*);
 - *zlecenie typu zrealizuj i anuluj* (ang. *fill and kill*);
 - *zlecenie typu zrealizuj lub anuluj* (ang. *kill or fill*);
- f) numer, datę i godzinę wystawienia zlecenia.

Zlecenie *dzienne* jest ważne w dniu wystawienia i może ono uczestniczyć w fazie notowań ciągłych i fazie notowań kursu jednolitego. Zlecenie typu: *ważne do końca notowań „instrumentu”* może być składane w dowolnej fazie aktualnej sesji. Może ono uczestniczyć zarówno w fazie notowań ciągłych, jak i notowań kursu jednolitego. Zlecenie typu: *do dnia* jest ważne do określonej daty. Zlecenie *czasowe* jest ważne w dniu przekazania na giełdę do określonego czasu i może uczestniczyć zarówno w fazie notowań ciągłych, jak i w fazie kursu jednolitego. Zlecenie *tylko na aukcję* może uczestniczyć jedynie w fazie handlu opartej na ustalaniu kursu jednolitego i tylko w jednej konkretnej aukcji. Oferty kupna i sprzedaży typu *zrealizuj i anuluj* uczestniczą w fazie notowań ciągłych i są ważne do momentu zawarcia pierwszej transakcji lub pierwszych transakcji, jeżeli są realizowane w kilku transakcjach równocześnie. Oferty niezrealizowane są anulowane. W tym przypadku nie jest konieczne podawanie limitu ceny. Zlecenia typu *zrealizuj lub anuluj* uczestniczą tylko w systemie notowań

ciągłych i są ważne do momentu zawarcia pierwszych transakcji, przy czym nie muszą być one realizowane w całości. Zlecenia typu zrealizuj i/lub anuluj nie są ujmowane w tabeli zleceń i w przypadku gdy transakcje nie zostaną zawarte są one usuwane. Zlecenie takie może zawierać dodatkową opcję aktywacji – tzw. funkcję Stop Loss, która zawiera warunek przy jakim poziomie kursu energii elektrycznej zlecenie pojawi się na rynku. Zlecenia mogą być modyfikowane przez zmianę oferowanej ilości energii elektrycznej oraz limitu ceny. Jeżeli podczas modyfikacji zmniejszany jest wolumen nie ulega zmianie czas złożenia zlecenia. W pozostałych przypadkach modyfikacji, tj. zwiększeniu wolumenu i zmianie ceny oferty, otrzymują nowy czas zlecenia.

W zupełnie inny sposób określa się specyfikację kontraktów terminowych na energię elektryczną (zob. [11]).

Mechanizmy ustalania ceny transakcji na rynku dnia następnego oraz w systemie notowań ciągłych realizowane są w następujący sposób.

a) Maksymalizowany jest wolumen obrotu.

b) Minimalizowana jest różnica między skumulowanym wolumenem energii elektrycznej w zleceniach kupna i sprzedaży możliwych do zrealizowania po określonym kursie.

W przypadku braku możliwości jednoznacznego określenia kursu RDN stosowany jest następujący algorytm. Jeżeli istnieje więcej niż jedna cena ustalona za pomocą powyższej metody oraz gdy różnica pomiędzy skumulowanym wolumenem kupna i skumulowanym wolumenem sprzedaży wynosi zero, wyznaczana jest średnia cena między skrajnymi cenami, tj. maksymalną ceną kupna i minimalną ceną sprzedaży. Jeżeli istnieją dwie takie ceny, kurs wyznaczany jest losowo spośród tych cen. Jeżeli różnica pomiędzy skumulowanym wolumenem kupna i skumulowanym wolumenem sprzedaży ma te same znaki dla każdej ceny, wyznaczany jest taki kurs, który jest bliższy zmiany znaku tej różnicy. W przypadku natomiast gdy różnica pomiędzy skumulowanym wolumenem kupna i skumulowanym wolumenem sprzedaży ma różne znaki dla każdej ceny, kurs wyznaczany jest losowo z dwóch cen: najwyższej ceny z dodatnią różnicą i najniższej ceny z ujemną różnicą (zob. [11], s. 9-10).

Zlecenia na rynku dnia następnego w systemie kursu jednolitego realizowane są wówczas, gdy cena sprzedaży określona w ofercie z limitem ceny jest niższa od ustalonego kursu energii elektrycznej, zaś cena kupna określona z limitem ceny jest wyższa od ustalonego kursu energii elektrycznej. Zlecenie, w którym cena sprzedaży jest wyższa zaś cena kupna jest niższa od kursu energii elektrycznej nie są realizowane. Zlecenia kupna i sprzedaży złożone z limitem ceny równym kursowi energii elektrycznej mogą zostać zrealizowane częściowo, w całości lub mogą nie zostać zrealizowane. O kolejności realizacji zleceń złożonych z limitem ceny równym kursowi energii elektrycznej decyduje czas przyjęcia zleceń do systemu informatycznego giełdy.

Transakcje w systemie notowań ciągłych zawierane są po kursie równym limitowi ceny, jeśli został podany w zleceniu oczekującym na realizację. W pierwszej kolejności realizowane są zlecenia o najwyższym limicie ceny w przypadku zleceń kupna i o najniższym limicie ceny w przypadku zleceń sprzedaży. W przypadku zleceń z równymi limitami ceny będą one realizowane według czasu przyjęcia zlecenia, tj. zlecenie przyjęte wcześniej zostanie zrealizowane w pierwszej kolejności. Niektóre zlecenia mogą być realizowane częściowo, przy czym każda transakcja dotyczy przynajmniej 1 MWh energii.

Można zauważyć, że reguły opisane powyżej są tożsame z regułami aukcji dwustronnej polegającej na tym, że oferty kupna i sprzedaży składane są z dwóch stron rynku. Dodatkowo widzimy, że handel na TGE ma charakter cykliczny.

3. AUKCYJNY MODEL HANDLU NA TGE

3.1. AUKCJE I ICH RODZAJE

Aukcja jest jedną z najstarszych form wymiany towarów i ma ona najczęściej charakter publicznej sprzedaży, w której nabywcą towaru jest kupujący oferujący najwyższą cenę. Biorąc pod uwagę różne modele aukcji nabywca niekoniecznie płaci tę najwyższą cenę. Przedmiotem handlu aukcyjnego są specyficzne i na ogół niejednorodne i niestandardowe towary, usługi oraz inne dobra. Unikalnymi towarami są, np. dzieła sztuki, antyki itp.; towary niejednorodne zwłaszcza pod względem jakości to chociażby ryby, tytoń, herbata itp. Istnieje kilka najbardziej rozpowszechnionych metod organizowania aukcji i przetargów. Najbardziej znaną jest aukcja angielska (ang. *English auction*), kojarzona zazwyczaj z licytacją, w ramach której uczestnicy zgłaszają oferty ustne, przy czym kolejne zgłoszenia opiewają na coraz to wyższe kwoty. Ta forma sprzedaży polega na wzajemnym przebijaniu ofert przez licytujących. Kończy się w chwili osiągnięcia takiej ceny, której żaden z uczestników nie jest w stanie podwyższyć. Wygrywa ten kupujący, którego oferta okazała się najwyższa. Jednakże ten rodzaj aukcji spełnia swoją rolę jedynie wówczas, gdy każdy uczestnik podwyższa swoje oferty co najwyżej do wysokości własnej wyceny licytowanego towaru. Wycena jest maksymalna cena, którą za obiekt skłonny jest zapłacić kupujący. Przekroczenie wyceny własnej grozi stratą. W efekcie stosowania strategii licytacji towaru poniżej własnej wyceny zwycięzca zapłaci cenę zbliżoną do ceny rynkowej.

Drugim, pod względem znaczenia, rodzajem konkursu ofert jest przetarg pisemny (ang. *sealed-bid auction*), w ramach którego potencjalni nabywcy składają propozycje ofert na piśmie, na przykład w zapieczętowanych kopertach. Sprzedający dokonuje wyboru oferty najlepszej, a kupujący płaci cenę, którą zaproponował. Jest to, tak zwana, aukcja pierwszej ceny (ang. *first price sealed-bid auction*). Zawartość optymalnej oferty składanej przez kupującego zależy od kilku elementów. Najważniejszym z nich jest właściwa wycena dobra będącego przedmiotem przetargu. Wystawione na sprzedaż dobro kupuje ten, kto złożył ofertę najwyższą i płaci tyle ile zaoferował. Problemem jest złożenie optymalnej oferty, a więc takiej, która nie przekroczy wyceny wystawionego do sprzedaży obiektu. Przedstawienie optymalnej oferty zależy od wielu czynników, przede wszystkim od natężenia konkurencji. Przygotowanie takiej oferty jest klasycznym przykładem podejmowania decyzji w warunkach niepewności.

Inną formą przetargu pisemnego jest aukcja drugiej ceny (ang. *second price sealed-bid auction*), gdzie wybierana jest oferta najlepsza, ale cena którą płaci kupujący jest drugą w kolejności od najwyższej. Ten rodzaj aukcji wymyślił w 1961 roku William Vickrey i opisał w pracy [12]. Wspomniana praca jest jedną z pierwszych, w której analizowane są i inne teoretyczne modele aukcji, w tym model aukcji asymetrycznej.

Kolejnym znanym rodzajem aukcji jest aukcja holenderka (ang. *Dutch auction*). W tym przypadku cena wystawianego na sprzedaż towaru jest sukcesywnie obniżana, aż do momentu, gdy znajdzie się nabywca. Zgodnie z regułami aukcji holenderskiej handlują, na przykład komisje. Nazwa: aukcja holenderska, pochodzi stąd, że w taki sposób sprzedawane są kwiaty w Holandii. Ten rodzaj alokacji dóbr wykorzystywany jest, zwłaszcza podczas sprzedaży szybko psujących się towarów.

Istnieje wiele kryteriów pozwalających na grupowanie różnych typów aukcji. I tak, ze względu na zasady dostępu aukcje można podzielić na zamknięte i publiczne. W aukcjach zamkniętych (ang. *non – public auctions*) uczestniczy ograniczona liczba osób zaproszonych. Aukcjach publiczne (ang. *public auctions*) są, z kolei, otwarte dla wszystkich zainteresowanych. Ze względu na formę realizacji i zawierania transakcji wyróżnia się aukcje ustne (ang. *oral auctions*) oraz aukcje pisemne (ang. *written auctions*). Podczas trwania aukcji ustnych oferty są anonosowane publicznie. Tak jest, np. w przypadku aukcji angielskiej lub holenderskiej. Aukcje pisemne zwane także niemym (ang. *silent auctions*) polegają na tym, że oferty są przekazywane licytatorom w formie pisemnej – aukcje pierwszej i drugiej ceny.

Teoria aukcji jest jedną z bardziej znanych dziedzin współczesnej teorii ekonomii. Reguły aukcyjne są w dalszym ciągu doskonalone i w związku z tym można obecnie wyróżnić wiele innych rodzajów aukcji. Oprócz wymienionych najbardziej znanymi są: aukcja udziałów (ang. *auction of shares*), aukcja dwustronna (ang. *double auction*) oraz aukcja wieloobiektowa (ang. *multiple-object auctions*). Aukcje wieloobiektowe stosuje się podczas sprzedaży wielu obiektów. Każdy stara się zmaksymalizować swoją funkcję użyteczności poprzez dobór właściwych towarów i trafne prognozowanie ich wycen. Dotyczy to zarówno sprzedających, jak i kupujących. Użyteczność ta jest funkcją zależną od indywidualnych wycen (ang. *private valuations*) poszczególnych uczestników aukcji. W ofercie kupna, oprócz proponowanej ceny, kupujący powinien uwzględnić ilość towaru, którą jest skłonny nabyć. Oferty takie są akceptowane bądź też nie, zgodnie z jasno określonymi procedurami. Jedną z takich przykładowych procedur polega na tym, że każdy z kupujących biorących udział w aukcji wieloobiektowej ogłasza, którym z wystawionych na sprzedaż obiektów jest zainteresowany. Prowadzący aukcję sonduje zapotrzebowanie, po czym, jeśli ocenia je zadowalająco podwyższa cenę każdego z towarów powiedzmy o jednostkę. Ponownie sonduje popyt, na przykład pytając kupców czy w dalszym ciągu zainteresowani są kupnem towarów po aktualnej cenie. Następnie ponownie ogłasza, które z towarów w dalszym ciągu będą licytowane. Znowu podwyższa cenę, po czym procedura się powtarza. Czyny tak aż do momentu, gdy nie będzie nikogo, kto będzie w stanie zapłacić więcej. W przypadku ustalania ceny transakcji zarówno na giełdach towarowych, jak i towarowo-pieniężnych wykorzystywane są reguły aukcji dwustronnych. Jest to rodzaj aukcji, który w ostatnich latach znalazł się w kręgu zainteresowań coraz większej liczby badaczy. Warto dodać, że wielu z nich uważa, iż reguły aukcyjne są tożsame z regułami gry, a zatem aukcję można traktować jako grę, zaś jej uczestników jako graczy, których celem jest najkorzystniejsze nabycie (lub też sprzedaż – w przypadku aukcji dwustronnej – określonego towaru).

3.2. AUKCJE DWUSTRONNE

Aukcja dwustronna charakteryzuje się tym, że oferty składane są z dwóch stron rynku – zarówno ze strony kupujących, jak i sprzedających. W sensie teorii gier zarówno kupujący, jak i sprzedający są graczami. Ten sposób handlu jest charakterystyczny dla giełd towarowych oraz niektórych rynków finansowych. Na podstawie składanych ofert ustalane są dwie funkcje: popytu i podaży. Ich „przecięcie” wyznacza cenę jednostkową towaru oraz jego ilość przeznaczoną do alokacji. Przykładem może być ustalanie, tak zwanej, ceny otwarcia wielu walorów na giełdach papierów wartościowych i towarów na podstawie bieżących propozycji kupna i ofert sprzedaży. W tym przypadku mamy do czynienia ze statyczną wersją aukcji dwustronnej. Wymiana może odbywać się cyklicznie, a transakcje mogą być prowadzone w określonych momentach czasu. Cykle mogą się powtarzać.

W praktyce jeszcze częściej wykorzystywana jest dynamiczna wersja aukcji dwustronnej. Jej reguły są stosowane wówczas, gdy oferty kupna i sprzedaży wpływają sekwencyjnie, a cena może być ustalana na bieżąco pomiędzy poszczególnymi kontrahentami. W tym przypadku, na ogół, potencjalni kupujący i sprzedający reprezentowani są przez brokerów, którzy licytują się wzajemnie w ich imieniu, w godzinach pracy giełdy. Transakcje zawierane są w momencie, gdy obie strony dogadają się.

Interesującym przykładem zastosowań aukcji dwustronnej jest rynek pracy. W tym przypadku oferty składają zarówno pracodawcy, jak i potencjalni pracownicy. Pracodawcy to kupujący kandydaci na pracowników to odpowiednio sprzedający. Z transakcją mamy oczywiście do czynienia wówczas, gdy obie strony dogadają się i kandydat na pracownika zostanie przyjęty na określonych warunkach do firmy. Ceną transakcji jest umowa oraz ewentualne bonusy wynikające z jej zawarcia, a z drugiej strony, zakres obowiązków wynikający z podjęcia określonej pracy.

Badania nad modelem aukcji dwustronnej prowadzone są od czterech dekad. Badał je, między innymi, laureat Nagrody Nobla z 2002 roku Vernon L. Smith. Interesowały go reakcje uczestników wymiany, na zasadach aukcji dwustronnych, na wszelkiego rodzaju bodźce zewnętrzne. Wraz ze współpracownikami przeprowadzał symulacje transakcji rynkowych [9], [10]. Kolejnym wartym odnotowania badaczem aukcji dwustronnych jest R. Wilson [14]. Podał warunki konieczne, ale niedostateczne, na istnienie strategii w równowadze dla aukcji dwustronnej traktowanej jako pewien rodzaj gry dynamicznej z niepełną informacją. Idea osiągnięcia równowagi zaproponowana przez Wilsona polegała, między innymi, na tym, że wyceny kupujących i sprzedających ustawia się w pewnym porządku, po czym „przyporządkowuje” się kupcowi z najwyższą wyceną sprzedającego z najniższym kosztem (wyceną). Podobne przyporządkowanie czynione jest w stosunku do pozostałych handlowców: kupujący, który złożył drugą w kolejności od najwyższej ofertę przyporządkowany jest sprzedającemu o drugim w kolejności od najniższych wycenie, itd. Tego typu uporządkowanie gwarantuje równowagę. Wilson zauważył, że wraz ze wzrostem liczby handlowców rośnie efektywność aukcji, a ceny, w tym przypadku, dążą do rzeczywistych cen rynkowych. Nie określił on jednak w jednoznaczny sposób strategii równowagi. Podał jedynie w sposób indukcyjny kilka kroków algorytmu, który mógłby być pomocny przy konstrukcji strategii licytacji.

Inni badacze również rozstrzygali problem ustalania ceny transakcji dla mechanizmu aukcji dwustronnej traktując ją jako grę z niekompletną informacją powtarzaną wielokrotnie. W 1998 roku Gjerstad i Dickhaut w pracy [1], doszli do wniosku, że w sytuacjach rzeczywistych cenę równowagi najlepiej ustala rynek. W tym samym roku ukazała się książka A. Sadrieha [8], w której autor zaprezentował interesujący model aukcji dwustronnej rozumianej jako gra. Model ten zostanie wykorzystany w dalszych rozważaniach. Obecnie aukcje traktowane są jako gry bayesowskie¹ [13].

3.3. TEORETYCZNY MODEL AUKCJI DWUSTRONNEJ ROZGRYWANEJ CYKLICZNIE

Przyjmuje się, że aukcja dwustronna jest pewną formą wymiany, w której uczestniczy wielu kupujących i sprzedających. Niech $I = \{1, \dots, n\}$ oznacza zbiór graczy – uczestników aukcji. Zbiór I dzielimy na dwa rozłączne podzbiory: n_B – elementowy zbiór kupujących I_B oraz n_S – elementowy zbiór sprzedających I_S ($I = I_B \cup I_S$; $n = n_B + n_S$). W przypadku TGE zarówno kupujący, jak i sprzedający powinni mieć status członka giełdy albo być reprezentowani przez brokerów.

Przez $i \in I_B$ ($i = 1, \dots, n_B$) oznaczmy i -tego kupującego, przez $j \in I_S$ ($j = n_B + 1, \dots, n_S$) j -tego sprzedającego. Załóżmy, że na aukcji handluje się $q > 0$ towarami (mogą to być również jednostki określonego towaru). Każdy z kupujących jest skłonny zapłacić maksymalnie v_{ik} jednostek monetarnych za jednostkę k -tego towaru ($k = 1, \dots, q$). Innymi słowy, v_{ik} jest wyceną i -tego kupującego za jednostkę towaru k . Każdemu sprzedającemu przyporządkowane są koszty jednostkowe, odpowiadające kwotom, które chcą uzyskać sprzedawcy za jednostkę poszczególnych dóbr c_{jk} ($j = n_B + 1, \dots, n_S$). W dalszym ciągu wielkości te będziemy nazywali wycenami sprzedawców. Wyceny kupujących i sprzedających są liczbami nieujemnymi.

Kupujący i -ty ($i = 1, \dots, n_B$) chcąc nabyć towar k -ty ($k = 1, \dots, q$) przedkłada ofertę b_{ik} , która powinna być niższa od jego wyceny. Sprzedający j -ty ($j = n_B + 1, \dots, n_S$) przedkłada ofertę sprzedaży a_{jk} . Załóżmy, że kupujący i sprzedający doszli do porozumienia i zawarto transakcję uzgadniając cenę jednostki towaru na p . Wypłatę kupującego wyrażamy za pomocą wzoru:

$$\pi_{ik} = v_{ik} - p \quad (i = 1, \dots, n_B; k = 1, \dots, q), \quad (1)$$

gdzie:

π_{ik} jest wypłatą i -tego kupującego nabywającego towar (udział) k -ty,

p oznacza cenę transakcji,

v_{ik} jest wyceną i -tego kupującego odnoszącą się do jednostki k -tego towaru.

¹ Gry bayesowskie, podobnie jak gry w tradycyjnym tego słowa znaczeniu, w uproszczeniu składają się z następujących elementów: graczy – w przypadku aukcji dwustronnych są to kupujący i sprzedający, reguł – sposoby licytacji – oraz wypłat. Gracze stosują określone, ale zgodne z regułami gry, strategie. W przypadku aukcji są to oferty. W odróżnieniu od gier tradycyjnych gry bayesowskie charakteryzują się tym, że wypłaty zależą nie tylko od strategii graczy, ale również nieznanego parametru, który można utożsamiać ze stanem natury (zob. [3], s. 23 i odnośniki do literatury w cytowanej pracy). Gry takie rozwiązują się w nieco odmienny sposób niż tradycyjne. Definiuje się typy graczy i dla tych typów rozwiązuje się grę (określa równowagę).

Wypłata jest równa różnicy pomiędzy wyceną i -tego kupującego odnoszącą się do k -tego towaru, a ceną transakcji. Wypłata sprzedającego jest równa różnicy pomiędzy ceną transakcji, a jego wyceną, co wyraża się za pomocą wzoru:

$$\pi_{jk} = p - c_{jh} (j = n_B + 1, \dots, n; k = 1, \dots, q). \quad (2)$$

W trakcie licytacji nie wszystkie oferty są jednak brane pod uwagę i w związku z tym należy jasno określić cechy tych, które możemy uznać za dobre.

Cechy dobrej oferty

1) Oferta nie powinna przynosić strat (ang. *no-loss*), co oznacza, że wypłaty zarówno kupujących, jak i sprzedających nie powinny być ujemne; kupujący składają oferty mniejsze lub równe od ich wycen; sprzedający odwrotnie – większe lub równe.

2) Oferta nie powinna stwarzać komplikacji (ang. *no-crossing*) – wypłaty powinny być większe od tych, które uczestnicy aukcji mogliby uzyskać na wolnym rynku bez konieczności brania udziału w aukcji; oferta kupującego powinna być mniejsza lub równa od aktualnej rynkowej ceny sprzedaży; oferta sprzedającego musi być większa lub równa od aktualnej rynkowej ceny kupna.

3) Oferta powinna zmniejszać rozpiętość pomiędzy cenami kupna i sprzedaży (ang. *ask-bid-spread-reduction*); oferty powinny być składane poniżej ceny rynkowej.

Wykorzystując opisany model został skonstruowany algorytm stochastyczny pozwalający na prognozowanie ceny transakcji dla aukcji dwustronnej ([3], [4]).

4. ALGORYTM PROGNOZOWANIA CENY TRANSAKCJI DLA AUKCJI DWUSTRONNEJ

4.1. ALGORYTM I OBLICZENIA

Zakładamy, że aukcja ma charakter cykliczny, a czas trwania cyklu jest zmienną losową. Zakładamy, że handlowcy przedkładają swoje oferty dotyczące obiektu k ($k = 1, \dots, q$) w trakcie trwania pojedynczego cyklu z , przy czym τ_z jest czasem trwania tego cyklu. Tak więc zgodnie z dotychczas przyjmowanymi oznaczeniami zakładamy, że i -ty ($i = 1, \dots, n_B$) kupujący składa ofertę $b_t(p_i)$ zaś j -ty ($j = n_B + 1, \dots, n_S$) sprzedający ofertę $a_t(p_j)$, gdzie p_i jest proponowaną ceną, b_t , a_t oznaczają liczbę jednostek towaru (jednostek energii), które zamierzają nabyć kupujący lub sprzedać sprzedający w momencie t . Oferta oprócz ceny zawiera liczbę jednostek towaru, który sprzedający zamierza sprzedać, kupujący zaś nabyć. Ich oferty są ważne jeżeli spełniają warunki tzw. dobrej oferty opisane powyżej. Dla potrzeb algorytmu, w dalszych rozważaniach wykorzystane zostaną następujące założenia.

Z1. Ceny p_i oraz p_j podane w ofercie spełniają następujący warunek:

$$p_i, p_j \in [\underline{p}, \bar{p}] \quad (3)$$

gdzie $\underline{p}$ jest minimalną ceną akceptowaną przez organizatorów danej aukcji, $\bar{p}$ zaś maksymalną.

Z2. Popyt przewyższa podaż, co formalnie zapisujemy jako:

$$\sum_{t=1}^{\tau_z} \sum_{i=1}^{n_B} b_t(p_i) 1_{\{p_i \in [\underline{p}, \bar{p}]\}} \leq \sum_{t=1}^{\tau_z} \sum_{j=n_B+1}^n a_t(p_j) 1_{\{p_j \in [\underline{p}, \bar{p}]\}}. \quad (4)$$

gdzie:

$1_{\{\cdot\}}$ jest indykátorem zdarzenia,

p_i jest ceną jednostki towaru (jednostki energii) proponowaną przez i -tego gracza ($i = 1, \dots, n$),

b_t jest liczbą jednostek towaru, które i -ty kupujący zamierza nabyć,

a_t jest liczbą jednostek towaru, które j -ty sprzedający zamierza sprzedać.

Stochastyczny algorytm wyznaczania ceny transakcji p^* .

1) Sprawdzamy, czy spełnione są warunki (3) i (4). Jeżeli spełnione nie są transakcja nie dochodzi do skutku.

2) Wybieramy maksymalną i minimalną ilość towaru, jaką zamierzają kupić/sprzedać gracze

$$c_{\min} = \min\{b_1, \dots, b_{n_B}, a_{n_B+1}, \dots, a_{n_B+n_S-1}, a_n\},$$

$$c_{\max} = \max\{b_1, \dots, b_{n_B}, a_{n_B+1}, \dots, a_{n_B+n_S-1}, a_n\},$$

$$b_i = \sum_{t=1}^{\tau_z} b_{it}, b_i \geq 0, i = 1, \dots, n_B,$$

$$a_j = \sum_{t=1}^{\tau_z} a_{jt}, a_j \geq 0, j = n_B + 1, \dots, (n_B + n_S - 1), n,$$

3) Wyznaczamy prostokąt $D \subset \mathbb{R}^2$ ($\mathbb{R}^2$ oznacza płaszczyznę), którego wierzchołkami są punkty $(c_{\min}, \underline{p})$, $(c_{\min}, \bar{p})$, $(c_{\max}, \underline{p})$, $(c_{\max}, \bar{p})$.

4) Za pomocą generatora liczb losowych generujemy losową liczbę punktów M_0 ($M_0 \leq 10000$) oraz liczbę iteracji k ($k \leq 100$). Punkty te powinny znajdować się wewnątrz prostokąta D , to znaczy $d_m(c_m, p_m) \in D$ ($m = 1, \dots, M_0$).

Punkty numerujemy w kolejności losowania.

a) Na początek przyjmujemy, $n := 1$. Dla każdego punktu d_m ($m = 1, \dots, M_{n-1}$) określamy „obszar przyciągania”

$$U(d_m, \varepsilon_n) = \sqrt{(\bar{p} - \underline{p})^2 + (c_{\max} - c_{\min})^2} \cdot \frac{\ln(n+1)}{n}. \quad (5)$$

Jeżeli w obszarze przyciągania znajduje się pewna liczba ofert kupna b_i i sprzedaży a_j , tj.

$$\exists_{i \in I_B} b_{it} \in U(d_m, \varepsilon_n) \quad (i = 1, \dots, n_B)$$

$$\exists_{j \in I_S} a_{jt} \in U(d_m, \varepsilon_n) \quad (j = n_B + 1, \dots, n)$$

to punkty te będą brane pod uwagę przy ustalaniu prognozowanej ceny transakcji. Zliczamy te punkty i przyjmujemy, że ich liczba wynosi M_n . Pozostałe punkty odrzucamy.

b) Przechodzimy do następnej iteracji $n := n + 1$. Punkty $d_m(c_m, p_m)$ ($m = 1, \dots, M_{n-1}$) „otaczamy” zbiorem $U(d_m, \varepsilon_n)$ o coraz mniejszym promieniu podobnie jak to miało miejsce w punkcie 4a.

c) Powtarzamy 4b do momentu gdy $M_n = 0$ jeżeli $n < k$ lub k razy jeżeli okaże się, że $M_n > 0$.

– Jeżeli $M_n = 1$ ($n < k$) to szukany punktem jest ten który pozostał: $d_{M_n}(c_{M_n}, p_{M_n})$, zaś ceną transakcji będzie druga współrzędna tego punktu $p^* = p_{M_n}$.

– Jeżeli $M_n > 1$ oraz wykonaliśmy k - iteracji (k - krotnie został zmniejszony promień zbioru przyciągania) to spośród M_n pozostałych punktów wybieramy ten, który podczas pierwszej iteracji zawierał w obszarze przyciągania największą liczbę ofert kupna i sprzedaży. Przyjmijmy, że jest to punkt $d_{m_0}(c_{m_0}, p_{m_0})$. W pierwszej kolejności pod uwagę brane są oferty, które w trakcie trwania cyklu, najwcześniej zostały przedłożone. Ceną równowagi będzie druga współrzędna tego punktu $p^* = p_{m_0}$.

Powyższy algorytm należy do klasy algorytmów stochastycznych i jego zaletą jest to, że nie wymaga on znajomości wycen graczy, ani też ich rozkładów tychże wycen. Jego idea opiera się na identyfikacji punktu na płaszczyźnie, którego jedną ze współrzędnych jest właśnie cena transakcji.

W tabeli 1 zostały pokazane wyniki przykładowej aukcji z dnia 12.01.2010 dla danych pochodzących z TGE (rynek dnia następnego). Tabela 1 zawiera rzeczywistą cenę transakcji oraz ceny teoretyczne wyliczoną za pomocą algorytmu oraz generatora liczb losowych, a także błędy obliczeń.

Zaprezentowana metoda prognozowania dla aukcji dwustronnej różni się od klasycznych metod prognozowania. Nie można bowiem powiedzieć, że szereg czasowy złożony z cen wygenerowanych podczas licytacji jest dobrze opisany i można go wykorzystać do budowy modelu ekonometrycznego, stanowiącego podstawę sporządzania prognozy. Heurystyczne metody prognozowania również zawodzą, ponieważ licytacja trwa na ogół krótko i bardzo często ma charakter dynamiczny. Ze względu na stochastyczny charakter testowanego algorytmu do porównań została wybrana metoda prognozowania ceny transakcji za pomocą generatora liczb losowych. Cena wybierana jest losowo z przedziału $[\underline{p}, \bar{p}]$, gdzie $\underline{p}$ jest ceną minimalną w konkretnym cyklu, zaś $\bar{p}$ jest ceną maksymalną.

Procentowe błędy prognozowanych cen transakcji dla jednej doby notowań z TGE przedstawiono w tabeli 1. Błędy liczone były w następujący sposób

$$\hat{\varepsilon} = \frac{|\hat{p}_i - p_i|}{p_i},$$

gdzie $\hat{p}_i$ oznacza prognozowaną cenę transakcji,
 p_i jest ceną rzeczywistą (obie ceny dotyczą i -tego cyklu).

Tabela 1

Ceny transakcji rzeczywiste i prognozowane obliczane za pomocą algorytmu oraz generatora liczb losowych dla danych pochodzących z TGE
(rynek dnia następnego) z dnia 12.01.2010

Cykl (godzina)	Średnia rzeczywista cena transakcji jednostki energii elektrycznej PLN/MWh	Prognozowana teoretyczna cena transakcji jednostki energii elektrycznej PLN/MWh	Prognozowana cena transakcji obliczona za pomocą generatora liczb losowych PLN/MWh	Procentowy względny błąd prognozy pomiędzy rzeczywistą ceną transakcji, a ceną transakcji obliczoną za pomocą algorytmu (w %)	Procentowy względny błąd prognozy pomiędzy rzeczywistą ceną transakcji, a ceną transakcji obliczoną za pomocą generatora liczb losowych (w %)
1	169,02	174	178,01	2,95	5,32
2	163,68	161	168,03	1,64	2,66
3	158,68	161	152,23	1,47	4,07
4	159,35	163	149,98	2,3	5,88
5	162,38	169	172,44	4,08	6,2
6	170,69	174	184,78	1,94	8,25
7	191,74	173	198,16	9,8	3,35
8	183,65	186	188,22	1,28	2,49
9	202,98	199	166,12	1,96	18,16
10	217,02	214	210,11	1,4	3,18
11	218,18	222	234,44	1,75	7,45
12	222,69	231	205,68	3,73	7,64
13	221,84	216	234,28	2,63	5,61
14	222,45	225	212,03	1,15	4,68
15	217,2	220	222,05	1,23	2,23

cd. tabeli 1

Cykl (godzina)	Średnia rzeczywista cena transakcji jednostki energii elektrycznej PLN\MWh	Prognoszowana teoretyczna cena transakcji jednostki energii elektrycznej PLN\MWh	Prognoszowana cena transakcji obliczona za pomocą generatora liczb losowych PLN\MWh	Procentowy względny błąd prognozy pomiędzy rzeczywistą ceną transakcji, a ceną transakcji obliczoną za pomocą algorytmu (w %)	Procentowy względny błąd prognozy pomiędzy rzeczywistą ceną transakcji, a ceną transakcji obliczoną za pomocą generatora liczb losowych (w %)
16	216,09	221	220,06	2,27	1,84
17	275,76	280	280,07	1,54	1,56
18	298,58	302	290,12	1,15	2,83
19	286,45	294	276,08	2,6	3,62
20	275,3	282	288,05	2,4	4,63
21	223,15	224	240,44	0,38	7,75
22	165,63	168	170,02	1,43	2,65
23	205,86	210	208,06	2,01	1,07
24	178,07	180	198,16	1,08	11,28

Źródło: opracowanie własne.

Średni procentowy błąd prognozy w przypadku algorytmu wynosi: 2,26, w przypadku generatora liczb losowych: 5,18. Różnica między wymienionymi wartościami jest znacząca.

W celu porównania jakości dwóch reguł prognozowania cen transakcji $\hat{p}_t$ oraz $\tilde{p}_t$ mające na celu stwierdzenie ich równoważności, ewentualnie istotnych różnic między nimi lub wskazanie lepszej ewentualnie gorszej, wymaga zastosowania testu równoważności procedur $\hat{p}_t$ i $\tilde{p}_t$. Hipotezę zerową o równoważności procedur $\hat{p}_t$ i $\tilde{p}_t$ możemy przedstawić w postaci

$$H_0: P(\hat{\varepsilon}_t > \tilde{\varepsilon}_t) = P(\hat{\varepsilon}_t < \tilde{\varepsilon}_t) = 0.5.$$

Hipotezę H_0 można testować przeciwko nierównoważności procedur

$$H_1: P(\hat{\varepsilon}_t > \tilde{\varepsilon}_t) \neq P(\hat{\varepsilon}_t < \tilde{\varepsilon}_t).$$

Hipotezą alternatywną może być również hipoteza $H_2: P(\hat{\varepsilon}_t > \tilde{\varepsilon}_t) < 0.5$, która mówi, że prognoza $\hat{p}_t$ jest lepsza od $\tilde{p}_t$, a także $H_3: P(\hat{\varepsilon}_t > \tilde{\varepsilon}_t) > 0.5$, która mówi, że prognoza $\hat{p}_t$ jest gorsza od $\tilde{p}_t$. Do testowania powyższych wariantów hipotez stosuje się test znaków, test Wilcoxona, test Walsha. W każdym z wymienionych testów zakłada się, że pary błędów $(\hat{\varepsilon}_1, \tilde{\varepsilon}_1)$, $(\hat{\varepsilon}_2, \tilde{\varepsilon}_2)$, ..., $(\hat{\varepsilon}_t, \tilde{\varepsilon}_t)$ tworzą ciąg niezależnych zmiennych losowych dwuwymiarowych o jednakowych rozkładach. Zaletą powyższych testów jest to, że można je stosować do bardzo małych prób.

Przeprowadzony zostanie test Wilcoxona ([6], s. 123). Ponownie do badań zostaną użyte dane zaprezentowane w tabeli 1. Porównane zostaną obliczenia wynikające ze stosowania algorytmu (x) oraz generatora liczb losowych (y). W tym celu procentowe błędy obliczeń dla obydwu metod zostały ustawione w kolejności od najmniejszego do największego:

0,38	1,07	1,08	1,15	1,15	1,23	1,28	1,4	1,43	1,47	1,54	1,56	1,64	1,75
x	y	x	x	x	x	x	x	x	x	x	y	x	x
1,84	1,94	1,96	2,01	2,23	2,27	2,3	2,4	2,49	2,6	2,63	2,65	2,66	2,83
y	x	x	x	y	x	x	x	y	x	x	y	y	y
2,95	3,18	3,35	3,62	3,73	4,07	4,08	4,63	4,68	5,32	5,61	5,88	6,2	7,45
x	y	y	y	x	y	x	y	y	y	y	y	y	y
7,64	7,75	8,25	9,8	11,28	18,16								
y	y	y	x	y	y								

Statystyka testowa U wynosi 479 (zob. [6]). Autorzy pracy [6] podają, że w przypadku dużych licznosci U ma rozkład $N\left(\frac{1}{2}n_1n_2, \sqrt{\frac{n_1n_2(n_1+n_2+1)}{12}}\right)$ przy czym przybliżenie jest dokładne gdy licznosci prób n_1 i n_2 spełniają $n_1, n_2 > 3$ i $n_1 + n_2 \geq 20$. W związku z tym można przyjąć, że $N(288; 95,05)$. Na poziomie istotności $\alpha = 0,05$ obszarem krytycznym jest: $\{0; 192,95\} \cup \{383,05; 576\}$. Obliczona statystyka $U = 479$

mieści się w tym przedziale, a zatem procedury $\hat{p}_t$ i $\tilde{p}_t$ nie są równoważne. Dodatkowo można wysnuć wniosek, że $\hat{p}_t$ jest lepsza od $\tilde{p}_t$, bo $\langle \hat{\varepsilon}_t \rangle < \langle \tilde{\varepsilon}_t \rangle$, gdzie: $\langle \cdot \rangle$ jest symbolem średniej.

4.2. WNIOSKI

Algorytm zaprezentowany w pracy jest algorytmem stochastycznym i jego niewątpliwą zaletą jest to, że nie wymaga on znajomości wycen licytowanego towaru dla poszczególnych uczestników aukcji ani też rozkładów tychże wycen. Jego idea polega na identyfikacji punktu na płaszczyźnie, przy czym jedną ze współrzędnych jest cena transakcji. W pracach [2], [3] oraz [7] był on testowany dla danych z WGT S.A. (Warszawska Giełda Towarowa). Rozważane były różne obszary przyciągania. Wyniki prezentowane we wszystkich cytowanych pracach, a także w niniejszej potwierdzają, że algorytm jest dobrym narzędziem służącym do identyfikacji ceny transakcji w przypadku aukcji dwustronnej. Algorytm jest zbiorem kroków, które kolejno wykonane prowadzą do identyfikacji takiego obszaru przyciągania, o możliwie najmniejszym promieniu, który zawiera jak najwięcej ofert i wyboru spośród tychże ofert najlepszej. Badania potwierdziły efektywność tego algorytmu.

Z drugiej strony zostało pokazane, że procedury handlu na TGE zostały tak opracowane, że pasują do reguł aukcji dwustronnej. Reguły aukcji dwustronnych stanowią bardzo interesujący mechanizm służący do alokacji różnorodnych dóbr i pieniędzy pośród wielu kupujących i sprzedających. Mechanizm ten generuje bowiem w miarę „sprawiedliwą” alokację dóbr i pieniędzy pomiędzy wszystkich uczestników konkretnego rynku. Reguły aukcji dwustronnych stosowane są w praktyce coraz częściej, głównie dzięki zainteresowaniu rynków kapitałowych.

5. DODATEK

Fragment programu napisanego w Pascalu (Delphi) ilustrujący obliczenia związane z tworzeniem obszarów przyciągania, w których zliczane będą oferty kupna i sprzedaży został zaprezentowany poniżej

```
begin
clrscr;
ile_buy:=zpliku(buy_plik,buy_tab); }
ile_sell:=zpliku(sell_plik,sell_tab);
for i:=1 to ile_sell do
begin
writeln(' ',i,' ',sell_tab[i,1]:2:0,' ',sell_tab[i,2]:2:2);
if i mod 23 = 0 then
end;
if (odp='n') OR (odp='N') then war13exit;
clrscr;
```

```

suma_buy:=war13pocz(ile_buy,buy_tab);
suma_sell:=war13pocz(ile_sell,sell_tab);
{if suma_buy>suma_sell then war13exit;}
ustaw_wierzcholki_prostokata;
max_dyst:=sqrt(sqr(p_max-p_0)+sqr(c_max-c_min));
for i:=1 to k do
begin
k_indeksy[i]:=exp((1+i/(i+1))*ln(1/i));
end;
obszar_przyciagania[1]:=(ln(k_indeksy[1]+1)/k_indeksy[1])*max_dyst;
for i:=2 to k do
begin
sumka:=0;
for j:=1 to i do sumka:=sumka+k_indeksy[j];
nawias:=(5/6-sumka/3);
obszar_przyciagania[i]:=max_dyst*(ln(i+1)/(i+1));
end; {obszary przyciagania obliczone.}

```

Politechnika Warszawska

LITERATURA

- [1] Dickhaut J., Gjerstad S., [1998], *Price formation in double auction*, „Games and Economic Behaviour”, 22, pp. 1-29.
- [2] Drabik E., [2007], *Aukcje w teorii i praktyce*, Wydawnictwo SGGW w Warszawie.
- [3] Drabik E., [2006], Eksperymentalne wyznaczanie ceny transakcji dla aukcji dwustronnej na podstawie danych z WGT S.A., „Przegląd Statystyczny”, 4 (53), 7-24.
- [4] Drabik E., [2000], *Zastosowania Teorii Gier do Inwestowania w Papiery Wartościowe*, Wydawnictwo Uniwersytetu w Białymstoku, Białystok.
- [5] Drabik E., [1999], *Aukcje jako gry*, „Przegląd Statystyczny”, 4 (46), 449-467.
- [6] Kryszicki K., Bartos J., Dyczka W., Królikowska K., Wasilewski M., [1994], *Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach. Cz. II. Statystyka matematyczna*, Wydawnictwo Naukowe PWN, Warszawa.
- [7] Majewska E., [2006], *Wykorzystanie reguł aukcyjnych do alokacji dóbr na przykładzie Warszawskiej Giełdy Towarowej*, praca doktorska, SGGW, Warszawa.
- [8] Sadrieh A., [1998], *The Alternating Double Auction Market. A Game Theoretic and Experimental Investigation*, Springer –Verlag, Berlin Heidelberg.
- [9] Smith V.L., [1982], *Microeconomic systems as experimental science*, „American Economic Review”, 72, 923-955.
- [10] Smith V.L., Williams A.W., Bratton W.K., Vannoni M.O., [1982], *Competitive market institutions: double auctions vs. sealed bid-offers auctions*, „The American Economic Review”, Vol. 72, No. 1, 58-77.
- [11] „Szczegółowe zasady obrotu i rozliczeń dla energii elektrycznej na Rynku Dnia Następnego z dnia 29.09.2009”, www.polpx.pl
- [12] Vickrey W., [1961], *Counterspeculation, auctions, and competitive sealed tenders*, „Journal of Finance”, 16, 8-37.
- [13] Wolfstetter E., [1996], *Auctions in introduction*, „Journal of Economic Surveys”, 10, 367-420.
- [14] Wilson R., [1992], *Strategic analysis of auctions*, [in:]: Aumann R.J., Hart S. (eds), *Handbook of Game Theory and Applications*, Vol. 1, 228-279, North- Holland, Amsterdam London New York Tokyo.

- [15] Xia M., Stallaert J., Whinston A.B., [2004], *Solving the combinatorial double auction problem*, „European Journal of Operational Research”, Article in press, www.sciencedirect.com, 1-13.
- [16] www.polpx.pl
- [17] www.tge.pl

Praca wpłynęła do redakcji w czerwcu 2010 r.

WYKORZYSTANIE REGUŁ AUKCYJNYCH DO HANDLU ENERGIĄ W POLSCE

Streszczenie

Rynek energii w Polsce nabrał charakteru konkurencyjnego w końcu lat 90. XX wieku. Wówczas to w większości krajów europejskich uchwalono nowe prawo (w Polsce w 1997 r.), które umożliwiło tworzenie wewnętrznych rynków energii elektrycznej. Pod koniec 1999 r. powstała w Polsce Towarowa Giełda Energii Elektrycznej (TGE), która niemalże od początku swego istnienia spełniała m.in. rolę mechanizmu rynkowego służącego do wyznaczania obiektywnej ceny energii.

Jako pierwszy na TGE uruchomiony został rynek „spot” (kasowy) energii elektrycznej, na którym ceny ustalane są na zasadzie reguł aukcji dwustronnej. Aukcja dwustronna jest taką formą wymiany towarów i pieniędzy, podczas której oferty składane są zarówno przez kupujących, jak i sprzedających. W każdej ofercie oprócz ceny podawana jest ilość towaru, którą kontrahenci zamierzają nabyć lub sprzedać.

Badania oraz eksperymenty symulacyjne dotyczące funkcjonowania aukcji dwustronnych prowadzone są na szeroka skalę. Zapoczątkował je w latach 60. XX wieku, późniejszy laureat Nagrody Nobla, Vernon L. Smith. Próby modelowania mechanizmów aukcji dwustronnej oraz zachowań jej uczestników podejmowało wielu autorów, takich jak: Wilson (1987), Friedman (1991), Easley i Ledyard (1993), Gjestad i Dichaut (1998), Sadierih (1998) oraz Xia, Stallaert, Whinston (2004). Celem pracy jest dopasowanie jednego z istniejących modeli aukcji dwustronnej do opisu funkcjonowania TGE. Do tego celu zostanie wykorzystany model Sadieriha, który dobrze oddaje reguły aukcyjne. Zostanie również zaprezentowany algorytm wyznaczania ceny transakcji dla aukcji dwustronnych (Drabik 1999), przy czym wyniki działania tego algorytmu zostaną porównane z rzeczywistymi cenami transakcji, które zostały odnotowane na TGE.

Słowa kluczowe: aukcja dwustronna, prognozowanie ceny transakcji, rynek energii

APPLICATION OF AUCTION RULES ON THE POLISH POWER EXCHANGE

Summary

The Polish energy market gained its competitive character in late 1990s. At that time in majority of European countries a new law was enacted (in Poland – in 1997), which enabled the creation of internal energy markets. The Polish Power Exchange has been functioning since the end of 1999. However, from the very onset it has constituted a vital component of under grounding liberalization of electricity market. Since it was created the Polish Power Exchange has served as a market mechanism for setting objective energy market price. Support and control of the Polish Financial Supervision Authority guarantee the security of concluded transactions.

The spot energy market was created as the first one and has functioned according to the rule of the double auction. Double auction are a form of money and goods exchange, during which the bids are made by purchases as well as by sellers. Besides the price, each bid discloses the amount of ware which is to be purchased by contracting parties. The rules of the double auctions, as well as of other auction types, are comparable to the rules of a game. The double auction is one of the most common exchange institution,

used extensively in stock markets, commodity markets and in markets for financial instruments, including options and futures.

Searches and simulation experiments pertaining to the functioning of the double auctions have been conducted on a wide scale. They were initiated in 60s in twenties centuries by the Vernon L. Smith. Many authors such as: Wilson (1987), Friedman (1991), Gode and Sunder (1993), Gjerstad and Dickhaut (1998), Sadrieh (1998), Xia, Stallaert, Whinston (2004) attempted to construe models of behavior of double auction participants. Although these models have furthered understanding of the interaction of individual behavior and institution in the double auction. The model of Sadrieh will be used for description of the auction rules applied to the spot energy trade on the Polish Power Exchange. Furthermore, an algorithm on the basis of which it is possible to forecast transaction prices is presented. The effectiveness of this algorithm will be compared with other traditional methods of forecasting transaction prices.

Key words: double auction, to forecast transaction price, energy market

MICHAŁ ŁUKOWSKI

ZASTOSOWANIE MODELU DRZEWA TRÓJMIANOWEGO DO WYCENY OPCJI MENEDŻERSKICH

1. WSTĘP

Rozwój polskich przedsiębiorstw wymusza efektywne zarządzanie kapitałem ludzkim. Znakomicie wyszkolona kadra menedżerów i specjalistów jest w tym procesie rozwoju niezbędna. W celu zatrzymania oraz motywowania pracowników stosowane są bardzo wyrafinowane sposoby motywacji. Za jeden z tych sposobów można uznać programy menedżerskie oparte na opcjach. Rosnąca popularność tego typu programów skłania do wnikliwej analizy instrumentów finansowych, na których są oparte.

Najważniejszymi zagadnieniami w przypadku opcji menedżerskich są odpowiednia konstrukcja warunków oraz właściwa wycena skonstruowanego programu motywacyjnego. Konstrukcja programu opcyjnego związana jest z zastosowaniem wielu warunków ograniczających, które mają niejako stanowić o skuteczności stworzonego systemu wynagradzania. Skutkiem tego jest wykorzystywanie w tego typu programach opcji egzotycznych, ale również konieczność modyfikowania standardowych modeli wyceny stosowanych w przypadku opcji. Modyfikacje mają służyć możliwości uwzględnienia charakteru warunków, na które mogą składać się okresy wyłączone z możliwości wykonania opcji, ekstrema osiąmane przez instrument bazowy, a także wskaźniki mówiące o możliwości wcześniejszego wykonania opcji przez menedżera. Warunki te muszą pozostawać w zgodzie z Międzynarodowym Standardem Sprawozdawczości Finansowej 2 [12]. Przedstawione w niniejszym artykule warunki w przekrojowy sposób pokazują charakter programów menedżerskich. Oczywiście, jest stosunkowo swobodna możliwość ich modyfikowania.

Modelami wyceny, które stosowane są w przypadku złożonych opcji, są modele wykorzystujące drzewa dwumianowe, trójmianowe, czy oparte na metodzie różnic skończonych. W artykule tym przedstawione jest podejście oparte na drzewie trójmianowym, które stanowi modyfikację rozwiązania zaproponowanego przez Hulla i White'a [8]. Na przykładzie tego modelu można w przystępny sposób przeprowadzić analizę opcji menedżerskich i sposób uwzględniania poszczególnych warunków w modelu wyceny. Zaproponowane modyfikacje stanowią rozszerzenie rozwiązania zaproponowanego przez Hulla i White'a polegające na rozwinięciu modelu z drzewa dwumianowego, do drzewa trójmianowego. Kolejna modyfikacja dotyczy wyników badań przeprowadzonych przez Huddarta i Langa [7] i polega na wprowadzeniu różnych wskaźników odejść dla okresu, w którym nabywane są uprawnienia (kiedy wykonanie opcji jest niemoż-

liwe) oraz dla okresu otwartego, kiedy przy spełnieniu dodatkowych warunków, opcja może zostać wykonana. W opracowanych modyfikacjach rozwinięty został również sposób obliczania wskaźnika odejść. Zamiast stałej stopy tego wskaźnika zaproponowana została postać funkcyjna, zależna od wartości akcji oraz czasu. Konstrukcja tej funkcji została oparta na rozwiązaniu zaproponowanym przez Musielę i Rutkowskiego [13]. Kolejna istotna modyfikacja polega na uchyleniu założenia, z którego wychodzili Hull i White, dotyczącego natychmiastowego wykonania opcji menedżerskiej w okresie otwartym, kiedy jest ona *in the money*. Stosowanie tego założenia powodowało zawyżanie wartości opcji i było sprzeczne z wynikami badań przeprowadzonych przez Carpenter [3] oraz Huddarta i Langa [7].

Przedstawiony w artykule model wyceny jest zgodny z założeniami MSSF 2 i pozwala na uwzględnienie wielu warunków ograniczających jednocześnie. Zostanie to zobrazowane zarówno w przypadku przedstawiania samej postaci modelu, jak i w przypadku przykładów obliczeniowych.

2. KONSTRUKCJA OPCJI MENEDŻERSKICH

Opcje menedżerskie, to opcje kupna wystawiane przez pracodawcę na akcje spółki, której program dotyczy, a nabywane przez pracownika, któremu program jest proponowany. Z uwagi na odmienny profil wypłaty od standardowych opcji kupna, będący wynikiem uwzględnienia warunków, od których może być uzależnione wykonanie opcji, podstawę konstrukcji opcji menedżerskich stanowią głównie opcje egzotyczne. Zyskuje się dzięki temu możliwość uwzględnienia w modelu wyceny ekstremów instrumentu bazowego, czy okresów wyłączonych z możliwości wykonania opcji.

Z uwagi na przedmiot niniejszego artykułu, interesująca jest grupa opcji wyodrębniona w książce pod redakcją Israela Nelkena przez Michaela Onga [16] – opcje egzotyczne uwzględniające stopień uwarunkowania ceny opcji od przebiegu ceny instrumentu bazowego (tzw. *path-dependence*).

Możliwość wykonania opcji menedżerskich zależy od czynników, które można podzielić na warunki rynkowe oraz nierynkowe przyznania uprawnień, a także warunki świadczenia usług. Warunki nabycia ustalone w dniu przyznania uprawnień (*grant date*) muszą zostać spełnione, aby możliwe było skorzystanie z programu przez osobę uprawnioną.

Warunki, które muszą zostać spełnione, są związane z działalnością gospodarczą jednostki, ale również bezpośrednio z osobą, której program motywacyjny jest oferowany. Warunki dotyczące osoby, nazywane są warunkami świadczenia usług i polegają na określeniu czasu, przez który pracownik musi świadczyć usługi na rzecz pracodawcy. Pozostałe warunki dzielą się na warunki rynkowe i warunki nierynkowe. Za rynkowe, uważa się te warunki, które leżą poza bezpośrednią kontrolą jednostki. Zalicza się do nich jeden z podstawowych warunków, czyli osiągnięcie przez akcje określonej ceny. Jak pokazują badania Lakonishoka i Lee [11], Seyhuna [19], [20], Narayanana i Seyhuna [14], pomimo interpretowania tych warunków jako leżących poza bezpośrednią kontrolą zainteresowanych programem osób, dzięki dostępowi do informacji niedostępnych dla wszystkich inwestorów, kadra zarządzająca jest często w stanie

przewidywać przyszłe ruchy cen akcji. Co więcej, w opracowaniu Herona i Liego [6], przedstawione zostały wyniki badań dotyczących nadnaturalnie niskiego poziomu cen akcji bezpośrednio przed wykonaniem opcji i nadnaturalnie wysokiego poziomu cen po ich wykonaniu. Wskazuje to jednoznacznie na możliwość wpływania menedżerów, w których interesie leży wykonanie opcji menedżerskich, na kształtowanie się cen akcji, szczególnie w okresie wykonywania opcji. Problem ten, poruszony w pracy Herona i Liego [6], wynika z nieuczciwej działalności menedżerów i tzw. kreatywnej księgowości zmierzającej do zwiększenia profitów osób objętych programem motywacyjnym. Poza osiągnięciem określonej ceny akcji do warunków rynkowych należy także osiągnięcie wewnętrznej wartości opcji (np. w przypadku opcji barierowej z barierą wejścia w dół i ceną wykonania w górę).

Za warunki nierynkowe uważa się te, które mogą być kontrolowane bezpośrednio przez jednostkę. Można do nich zaliczyć wszystkie sprawy związane z wewnętrzną działalnością firmy, a ich ustalenie może mieć dowolny charakter. Przy wycenie, spełnienie tych warunków przyjmować będzie postać zmiennej binarnej i w ten sposób decydować, czy opcja może zostać wykonana. W opracowaniu Preussa i Ślusarskiego [17] wskazano także na osiągnięcie celów indywidualnych, co w praktyce oznacza, że wszystko to, na co zainteresowana programem osoba ma wpływ, co nie jest poza jej wyraźną kontrolą, może być zaliczone do warunków nierynkowych nabywania uprawnień.

Ważnym elementem konstrukcyjnym opcji menedżerskich jest brak możliwości przeniesienia praw przysługujących osobie uprawnionej na osoby trzecie oraz ograniczenie możliwości wpływania menedżerów na wynik finansowy poprzez wyłączenie możliwości wykonania opcji w okresie przypadającym na publikację sprawozdań finansowych, co prowadzi do modyfikacji okresu wykonania opcji. Specyfikacja opcji menedżerskich powinna w swojej konstrukcji uwzględniać także prowadzoną przez przedsiębiorstwo politykę dywidend, która mogłaby prowadzić do wykorzystywania spadku cen akcji do osiągnięcia dodatkowych zysków wynikających z programu motywacyjnego.

Programy menedżerskie, ze względu na sposób ich rozliczania, można podzielić na rozliczane w środkach pieniężnych [6] oraz rozliczane w instrumentach kapitałowych. Rozliczanie w instrumentach kapitałowych wiąże się dla przedsiębiorstwa z koniecznością emisji akcji bądź skupowaniem istniejących akcji. Zgodnie z MSSF 2 odmienny jest także sposób ujmowania tych programów – jako zobowiązania, dług lub kapitał własny [9]. Ze względu na obszerność dyrektywy nie zostanie ona omówiona szczegółowo w niniejszym artykule. Warto jednak nadmienić, że dokładna analiza standardu rachunkowości stanowi niepodważalnie punkt wyjścia przy konstruowaniu modeli wyceny.

Opcje menedżerskie rozliczane w instrumentach kapitałowych są ujmowane natychmiast, jeżeli żadne dodatkowe warunki nie muszą zostać spełnione. W przypadku występowania takich warunków, transakcja musi być ujmowana w sprawozdawczości finansowej przez cały okres spełniania tych warunków. Wycena programu oparta jest w takim przypadku na wartości godziwej instrumentów, na które opcje są wystawione, na dzień przyznania, co w obliczu konieczności uwzględnienia warunków rynkowych i terminów danego programu oraz braku możliwości zdobycia adekwatnych danych rynkowych, prowadzi do wykorzystania teoretycznych modeli wyceny. Po ustaleniu war-

tości godziwej, wpływ na kontrakt mają już tylko warunki nierynkowe, które wpływają na liczbę instrumentów, których dany program będzie dotyczył. Programy rozliczane w środkach pieniężnych ujmowane są jako koszt i wyceniane na podstawie wartości godziwej zobowiązań na każdy dzień sprawozdawczy, wpływając na wyniki sprawozdawcze przedsiębiorstwa w danym okresie rozliczeniowym. Programy rozliczane w środkach pieniężnych posiadają wiele cech wspólnych z programami *Phantom Share Plan*, dających pracownikowi prawo otrzymania premii przy wzroście ceny akcji.

Poszukiwanie modeli wyceny uwzględniających charakterystykę opcji menedżerskich jest istotne ze względu na fakt, że standardowe modele wyceny opcji nie uwzględniają takich właściwości opcji menedżerskich [8], jak: okres nabywania uprawnień, utrata możliwości wykonania opcji w przypadku rezygnacji pracownika z zatrudnienia w okresie nabywania uprawnień, utrata możliwości wykonania opcji *out-of-money*, faktu, że pracownikowi nie wolno zbywać opcji menedżerskich oraz występowanie efektu rozmycia w przypadku emisji nowych akcji w celu rozliczenia programu menedżerskiego. Poza tym, jak w swojej pracy dowodzi Rubinstein [18], wartości otrzymywane przy wykorzystaniu standardowych modeli wyceny prowadzą do przeszacowania wartości opcji.

Właściwe określenie wartości godziwej opcji menedżerskich jest istotne nie tylko z punktu widzenia wymogów rachunkowości, ale również z punktu widzenia oczekiwań pracowników i udziałowców. Rozbieżność między wartością opcji dla przedsiębiorstwa i otrzymującego opcję wynika z faktu, że pracownik jest pozbawiony możliwości zbycia opcji lub jej zabezpieczenia, gdyż krótkie pozycje w opcjach firmy są niedozwolone [4]. Firma natomiast może zbyć swoje zobowiązanie lub zabezpieczyć swoją pozycję. Wartość opcji zależy jednak w dużej mierze od polityki wykonywania opcji przez menedżera, którego zachowanie, dzięki posiadaniu danych niedostępnych dla rynku, może różnić się od optymalnego zachowania inwestorów.

Sposób wyceny opcji menedżerskich w dużej mierze zależy od warunków, którymi program menedżerski został obarczony. W myśl MSSF 2, przy wycenie wartości godziwej opcji menedżerskich należy uwzględnić wszystkie warunki rynkowe, które wzięłyby pod uwagę dobrze poinformowane strony tej transakcji, a wycena powinna być prowadzona z punktu widzenia przedsiębiorstwa. Na wartość godziwą nie mają natomiast wpływu warunki nierynkowe, stąd nie bierze się ich pod uwagę przy wycenie (powodują one jedynie korektę liczby instrumentów objętych programem), a w przypadku ich niespełnienia, odwraca się rozpoznane koszty związane z programem motywacyjnym.

Jednym z czynników, który różni opcje menedżerskie od innych opcji i stanowi o konieczności modyfikowania modeli wyceny, jest wskaźnik odejść. Jest to obserwowalna wielkość, będąca stopą rezygnacji pracowników z zatrudnienia w danym przedsiębiorstwie. Istotne jest zarówno, kiedy dany pracownik objęty programem motywacyjnym rezygnuje z zatrudnienia (w okresie nabywania uprawnień, czy po jego upływie) oraz jakie stanowisko zajmuje. Jak w swoich badaniach przedstawili Huddart i Lang [7], miejsce w pracowniczej hierarchii ma wpływ na wysokość wskaźnika odejść. Pracownicy wyższych szczebli dłużej wstrzymują się od wykonania opcji, co związane jest z poziomem zamożności [7] oraz płynnością finansową [9]. Zbliżając wartość opcji menedżerskich do rzeczywistości można uznać, że wskaźnik odejść jest

zależny od czasu pozostającego do wygaśnięcia opcji oraz wartości akcji, na których dany program został oparty [15], zastępując tym samym stałą wartość tego wskaźnika w czasie.

Na wycenę programów motywacyjnych istotny wpływ ma konstrukcja opcji związana zarówno z czasem jej życia, jak i okresem nabywania uprawnień. Średni czas trwania programów motywacyjnych wynosi od ośmiu do dziesięciu lat [7], [9] i jest ściśle związany z celem, któremu program motywacyjny ma służyć. Zbyt długi okres, ze względu na dużą niepewność związaną z kształtowaniem procesów gospodarczych, może nie stanowić odpowiedniej motywacji. Zdecydowana większość programów daje możliwość wykonania opcji przed terminem wygaśnięcia (po okresie nabywania uprawnień). Jak pokazują badania Carpenter [3], menedżerowie korzystają z tej możliwości, gdyż średni czas wykonania opcji wynosi 5,83 roku, przy czym wszystkie badane firmy określiły czas trwania programów na 10 lat, natomiast w przypadku badań Huddarta i Langa [7] średni czas wykonania opcji wynosił 3,4 roku. Z punktu widzenia zarówno motywacyjnego, jak i modeli wyceny, istotny jest również okres nabywania uprawnień (z badań Carpenter [3] – 1,96 roku). W okresie tym konieczne jest spełnienie wszystkich ustalonych warunków, aby wykonanie opcji stało się możliwe w okresie otwartym (do momentu zapadalności). Uwzględnienie okresu nabywania uprawnień w procesie wyceny zbliża postać opcji menedżerskich do opcji bermudzkich.

Możliwość wcześniejszego wykonania opcji menedżerskich prowadzi do upodabniania ich do opcji barierowych [8]. Wysokość bariery może zostać narzucona przez wystawiającego opcję lub obliczona na podstawie historycznych danych dotyczących średniej ceny akcji w momencie wcześniejszego wykonywania opcji przez menedżerów. Historyczne dane powinny być dobrane do grupy pracowników, których program dotyczy i nie powinny obejmować okresu bezpośrednio po okresie nabywania uprawnień. W swoich badaniach nad wielkością tego wskaźnika, Huddart i Lang [7], rozpatrzyli przypadki pięćdziesięciu tysięcy pracowników z ośmiu firm w latach 1982-1994 i otrzymali wysokość bariery (wskaźnika wielokrotności) na poziomie 2,2 razy wyższym od ceny wykonania. Natomiast w badaniach Carpenter [3] dotyczących najwyższych rangą menedżerów z czterdziestu firm w latach 1979-1994, obliczono ten wskaźnik na poziomie 2,8. W przypadku kiedy ograniczenie nie zostało narzucone przez emitenta, oszacowanie wielkości współczynnika wielokrotności, jest niejako informacją dla przedsiębiorstwa, o polityce wykonywania opcji przez menedżerów.

Czynnikiem, którego właściwe oszacowanie ma znaczący wpływ na wycenę wszystkich opcji, w tym także opcji menedżerskich, jest współczynnik zmienności. Szacowanie zmienności jest bardzo szerokim zagadnieniem, wobec czego, ze względu na charakter niniejszego opracowania nie zostanie przedstawione. Szacując współczynniki zmienności na potrzeby badań wykorzystano modele z rodziny GARCH.

Przy wycenie programów menedżerskich warto zastanowić się nad efektem rozmycia, którego wpływ został w przedstawianym modelu wyceny zignorowany, na podstawie tezy wysuniętej przez Hulla i White'a [8], mówiącej, że wielkość emisji nowych akcji jest mała w stosunku do wszystkich wyemitowanych przez spółkę akcji.

3. MODEL WYKORZYSTUJĄCY DRZEWO TRÓJMIANOWE

Zadanie stworzenia modelu wyceny opcji menedżerskich pozwalającego na jednoczesne uwzględnienie zarówno wskaźnika odejść, jak i okresów wyłączonych z możliwości wykonania opcji, sprowadza problem do wyceny opcji bermudzkiej uwzględniającej w swojej konstrukcji barierę. Do wyceny opcji barierowych oraz bermudzkich wygodnym narzędziem jest model drzewa trójmianowego. Jest on znacznie prostszy w implementacji od modelu różnic skończonych, a przedstawienie jego konstrukcji pozwala na wyróżnienie poszczególnych modyfikacji wynikających z uwzględnianych w wycenie warunków programu w przystępny sposób.

Model wyceny opcji menedżerskich zaproponowany przez Hulla i White'a oparty został na modelu wyceny za pomocą drzewa dwumianowego. Jak jednak zaznaczają sami autorzy, do wyceny opcji (w tym opcji barierowych, do których w tym punkcie sprowadzą się opcje menedżerskie) znacznie efektywniejsze jest wykorzystanie drzewa trójmianowego. Dzięki zagęszczeniu poszczególnych gałęzi drzewa otrzymujemy węzły, w których wartości akcji odpowiadają określonym barierom. Dlatego zmodyfikowane zostało podejście Hulla i White'a, poprzez zastosowanie modelu drzewa trójmianowego.

Model omawiany w tym punkcie w przejrzysty sposób bierze pod uwagę możliwość oraz politykę wcześniejszego wykonania opcji (po okresie nabywania uprawnień) przez menedżera. Wcześniejsze wykonanie opcji wykorzystuje oszacowaną w punkcie dotyczącym konstrukcji wskaźnika odejść wielkość zarówno dla okresu nabywania uprawnień, jak i dla okresu otwartego. Postać jaką przyjmuje przedstawia wzór:

$$e = \ln(1 + u), \quad (1.1)$$

gdzie:

e – wskaźnik odejść stosowany w modelu,

u – wskaźnik odejść oszacowany na podstawie badań empirycznych.

W prezentowanym w artykule modelu zdecydowano się na modyfikację założeń poczynionych przez Hulla i White'a zakładających wielkość wskaźnika e na poziomie 0 w okresie nabywania uprawnień. Powodem wprowadzenia dwóch różnych wartości wskaźnika odejść są wyniki badań przeprowadzonych przez Huddarta i Langa [7] mówiące o odmiennym zachowaniu pracowników w okresie otwartym i zamkniętym. Poza tym, zgodnie z opracowaniem Noreena i Wolfsona [15], zmienność wskaźnika odejść w czasie, skłania do przedstawienia go za pomocą funkcji czasu i wartości akcji, zamiast, jak miało to miejsce w modelu Hulla i White'a, za pomocą stałej wartości.

Polityka wcześniejszego wykonywania opcji jest związana z wcześniej omawianym współczynnikiem wielokrotności, oznaczonym w przypadku tego opracowania przez M . Współczynnik ten ogranicza możliwość wykonania opcji przez ich posiadaczy (gdy została narzucona przez emitenta np. w postaci bariery) w okresie otwartym do momentu wygaśnięcia opcji (jednak nie w samym momencie wygasania opcji). Stanowi on dodatkowe zabezpieczenie przed wykonaniem opcji, przez co zwiększa skuteczność przywiązania menedżera do pracodawcy w okresie trwania programu

motywacyjnego. Z perspektywy analizy, istotny jest wpływ tego czynnika na sam sposób wyceny opcji menedżerskich, dlatego sposób oszacowania wartości tego parametru nie zostanie przedstawiony. Otóż, wprowadzenie go do moich rozważań przybliża postać opcji menedżerskiej do opcji barierowej. Niezależnie od tego, czy bariera została określona przez emitującego opcję, czy pozostaje tylko obserwacją polityki wykonania opcji stosowaną przez menedżerów, warto wziąć pod uwagę wpływ tego parametru na wartość opcji.

Wykorzystanie wskaźnika wielokrotności powoduje także ominięcie nieprawidłowości wyceny za pomocą modeli, w których za dane wejściowe należy uznać maksymalny okres, jaki może minąć od daty przyznania do terminu wygaśnięcia (określone w standardzie MSSF 2 jako oczekiwany czas życia opcji). Jest to związane z określeniem o ile wartość ceny akcji musi przekraczać cenę wykonania, aby możliwe było wcześniejsze wykonanie opcji [8]. Dzięki wykorzystaniu wskaźnika wielokrotności ominięta zostaje konieczność oszacowania oczekiwanej długości życia opcji (w myśl standardu MSSF 2).

4. POSTAĆ MODELU

Zmodyfikowanie oryginalnego modelu do postaci drzewa trójmianowego spowodowało konieczność zmiany wszystkich formuł wykorzystywanych w modelu wyjściowym. Ciekawym rozwiązaniem byłoby oparcie modelu na implikowanym drzewie trójmianowym zaproponowanym przez Dermana i Kanię [5], jednak ze względu na ograniczoną dostępność danych związanych z cenami opcji menedżerskich, niniejszy artykuł zostanie oparty na drzewie trójmianowym.

Zasady wyznaczania wartości opcji w każdym węźle [8] określają możliwość wykonania opcji tylko po okresie nabywania uprawnień i tylko w przypadku, jeżeli cena akcji przekracza cenę wykonania M razy (gdzie $M > 0$). Poza tym istnieje prawdopodobieństwo, że opcja zostanie utracona w czasie okresu nabywania uprawnień i jest ono równe $e\delta t$ dla każdego krótkiego okresu pomiędzy poszczególnymi węzłami w drzewie (gdzie: $e = \ln(1 + u)$, natomiast δt oznacza odstęp czasowy między kolejnymi węzłami) oraz istnieje prawdopodobieństwo, że opcja zostanie utracona w każdym małym czasie δt i wynosi ono $e\delta t$.

Zgodnie z założeniami oryginalnego modelu, jeżeli opcja ma zostać utracona po okresie nabywania uprawnień, to jeżeli jest *out of the money*, to faktycznie zostaje utracona, natomiast jeżeli jest *in the money*, to zostaje natychmiast wykonana. W przypadku przedstawionych badań postanowiono zmodyfikować to podejście. Dlatego kiedy opcja jest *in the money* oraz jesteśmy w fazie okresu otwartego, to pomimo możliwości wykonania opcji, nie jest to konieczne. Dlatego zaproponowany przez Hulla i White'a został zmodyfikowany. Modyfikacja polegała na zdyskontowaniu wartości otrzymywanych ze wzoru 1.6.

Wartości akcji są w każdym następnym węźle drzewa:

– w przypadku wzrostu ceny akcji w następnym węźle równe S_u , gdzie $S_u = S_0 u$,

$$u = e^{\sigma\sqrt{2\delta t}},$$

– w przypadku spadku ceny wartość akcji w węźle równe S_d , gdzie $S_d = S_0d$,
 $d = e^{-\sigma\sqrt{2\delta t}}$,

– w przypadku pozostania ceny na dotychczasowym poziomie równe S_m , gdzie
 $S_m = S_0m$, $m = 1$,

gdzie przez S_0 oznaczyłem wartość akcji w węźle wyjściowym.

Prawdopodobieństwo zmiany ceny opcji w każdym węźle wynosi:

– prawdopodobieństwo, że cena opcji wzrośnie, wynosi w każdym węźle:

$$p_u = \left(\frac{e^{r\delta t/2} - e^{-\sigma\sqrt{\delta t/2}}}{e^{\sigma\sqrt{\delta t/2}} - e^{-\sigma\sqrt{\delta t/2}}} \right)^2, \quad (1.2)$$

– prawdopodobieństwo, że cena opcji spadnie, wynosi w każdym węźle:

$$p_d = \left(\frac{e^{\sigma\sqrt{\delta t/2}} - e^{r\delta t/2}}{e^{\sigma\sqrt{\delta t/2}} - e^{-\sigma\sqrt{\delta t/2}}} \right)^2, \quad (1.3)$$

– prawdopodobieństwo, że cena pozostanie na niezmiennym poziomie wynosi:

$$p_m = 1 - p_u - p_d, \quad (1.4)$$

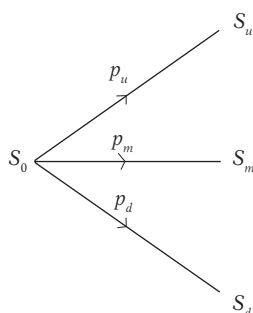
gdzie:

σ – współczynnik zmienności ceny akcji,

r – stopa wolna od ryzyka,

δt – długość kroku czasowego (odstęp czasowy między węzłami).

Proces budowania trójmianowego drzewa dla ceny akcji obrazuje rysunek 1.1:



Rysunek 1.1 Schematyczne przedstawienie drzewa trójmianowego ceny akcji

Konstrukcja drzewa trójmianowego dla ceny opcji, w przypadku gdy rozważamy N kroków o długości δt każdy, wygląda przy wykorzystaniu indukcji wstecznej dla konstrukcji opartej na drzewie dwumianowym (na podstawie artykułu Hulla i White'a [8]) następująco (z wprowadzonymi odpowiednimi zmianami (wzory 1.6-1.9).

Wartość opcji w węzłach końcowych:

$$f_{N,j} = \max(S_{N,j} - K, 0), \quad (1.5)$$

gdzie:

$f_{N,j}$ – wartość opcji w ostatnim węźle drzewa,

$S_{N,j}$ – wartość akcji w ostatnim węźle drzewa (odpowiadającym węzłowi opcji),

K – cena wykonania opcji.

W przedziale czasowym od momentu przyznania opcji do przedostatniego węzła $0 \leq i \leq N - 1$ wartość opcji wynosi:

– w przypadku kiedy minął okres nabywania uprawnień:

dla $i\delta t > \nu$ i $S_{i,j} \geq KM$ opcja przyjmuje wartość:

$$f_{i,j} = S_{i,j} - K, \quad (1.6)$$

dla $i\delta t > \nu$ i $S_{i,j} < KM$ opcja przyjmuje wartość:

$$f_{i,j} = (1 - e\delta t)e^{-r\delta t} [p_u f_{i+1,j+1} + p_m f_{i+1,j} + p_d f_{i+1,j-1}] + e\delta t \max(S_{i,j} - K, 0), \quad (1.7)$$

– w przypadku kiedy okres nabywania uprawnień nie minął:

$i\delta t < \nu$ opcja przyjmuje wartość:

$$f_{i,j} = (1 - e\delta t)e^{-r\delta t} [p_u f_{i+1,j+1} + p_m f_{i+1,j} + p_d f_{i+1,j-1}], \quad (1.8)$$

– wartość opcji w węźle początkowym drzewa:

$$f_{00}. \quad (1.9)$$

gdzie:

i – numer kroku czasowego,

ν – okres nabywania uprawnień,

δt – długość kroku czasowego,

$S_{i,j}$ – wartość akcji w węźle i, j ,

K – cena wykonania opcji,

M – wskaźnik wielokrotności,

$f_{i,j}$ – wartość opcji w węźle i, j ,

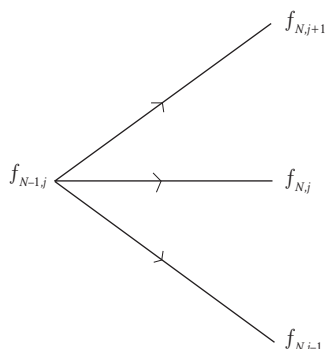
$f_{i+1,j}$ – wartość opcji w węźle $i + 1$, dla wartości ceny opcji na niezmiennym poziomie,

$f_{i+1,j+1}$ – wartość opcji w węźle $i + 1$, dla wartości ceny opcji w przypadku wzrostu,

$f_{i+1,j-1}$ – wartość opcji w węźle $i + 1$, dla wartości ceny opcji w przypadku spadku.

Istotną w tym momencie jest wprowadzona modyfikacja wzoru (1.6). Otóż postać zaproponowana przez Hulla i White'a nie uwzględnia dyskontowania wartości ze wzoru (1.6), co w mojej ocenie prowadzi do zawyżania wartości opcji. Jako, że każdy węzeł drzewa niejako dyskontuje wartości pobierane z poprzedniego węzła, dlatego w niniejszym opracowaniu zdecydowano się na wprowadzenie do wzoru (1.6) członu następującej postaci $e^{-r\delta t^*m}$, przez który wartość (1.6) jest w każdym węźle przemnożona. W członie tym występuje wartość m , która wskazuje przez ile okresów dana wartość powinna zostać przemnożona (czyli o ile okresów od ostatniego węzła drzewa powinna dana wielkość zostać zdyskontowana). Dla węzła N wartość m wynosi 0, dla węzła $N-1$ wartość ta wynosi 1, dla węzła $N-2$ wartość ta wynosi 2, itd. W ten sposób dyskontowane są wszystkie wartości opcji w węzłach, w których wykonanie opcji jest możliwe (czego moim zdaniem, w modelu wyjściowym niesłusznie brakowało). Uwzględnienie członów tej postaci ($e^{-r\delta t^*m}$) wynika z możliwości wykonania opcji w danym momencie (okres otwarty, cena akcji wyższa od ceny wykonania przemnożonej przez wskaźnik odejścia), jednak pracownik w zależności od swojej decyzji, może nie rezygnować z programu. W takim wypadku węzeł ten nie będzie ostatnim węzłem wyceny i stąd za zasadne uważam zdyskontowanie występującej w nim wartości opcji. Przez wprowadzoną modyfikację uchylone zostało założenie modelu wyjściowego polegające na konieczności natychmiastowego wykonania opcji, kiedy jest to możliwe.

Proces budowania trójmianowego drzewa dla wartości opcji obrazuje rysunek 1.2:



Rysunek 1.2 Schematyczne przedstawienie tworzenia drzewa trójmianowego ceny opcji

Rozszerzenie modelu zaproponowanego przez Hulla i White'a o metodę opartą na drzewie trójmianowym pozwala na zagęszczenie ilości węzłów, co w efekcie prowadzi do wzrostu prawdopodobieństwa trafienia na cenę akcji odpowiadającą ustalonej w programie barierze.

5. KONSTRUKCJA WSKAŹNIKA ODEJŚĆ

Konstrukcja wskaźnika odejść oparta została na publikacji Musieli i Rutkowskiego [13]. Funkcja $\lambda(S, t)$ interpretowana jest jako roczna stopa wskaźnika odejść spo-

wodowanych przedterminowym (przed terminem wygaśnięcia, po okresie nabywania uprawnień) wykonaniem opcji, które w momencie wykonywania są *in the money*. Parametr ten, jak zauważa Graeme West [21], może zostać oszacowany na podstawie danych historycznych dotyczących danego przedsiębiorstwa. Jednak w przypadku przeprowadzonego badania, możliwość pozyskania takich danych okazała się bardzo ograniczona. Poza tym celem niniejszego artykułu było przedstawienie powyższego parametru za pomocą funkcji kilku zmiennych. Stąd na podstawie opracowań Ayache [1], [2] czynniki λ przedstawione zostaną jako funkcje zależne od wartości akcji oraz czasu.

Własności funkcji $\lambda(S, t)$ [21]:

- $\lambda(S, t) = 0$, dla $S < K$ oraz jeżeli jest to okres zamknięty,
- $\lambda(S, t)$ rośnie wraz ze wzrostem ceny akcji S oraz czasu t oraz spadkiem współczynnika zmienności σ ,
- w okresach w których funkcja $\lambda(S, t) = 0$, funkcja $v(S, t) \neq 0$, co pozwala stwierdzić, że są to funkcje wzajemnie komplementarne.

W celu oszacowania funkcji $\lambda(S, t)$ wykorzystana zostanie publikacja Musieli i Rutkowskiego [13]. Autorzy zaproponowali, aby prawdopodobieństwo wcześniejszego wykonania opcji, które w mojej pracy będzie interpretowane jako wskaźnik odejścia, szacować na podstawie opcji barierowej. Bariera jest przy tym ustalana na podstawie danych historycznych dotyczących odejść – jest to wartość opisywana w badaniach Carpenter [3] oraz Huddarta i Langa [7], a także wspomniany przy okazji omawiania modelu Hulla i White’a współczynnik wielokrotności. Wzór zaproponowany przez Musielę i Rutkowskiego (1.10) odnosi się do okresu otwartego, w którym cena akcji jest wyższa od ceny wykonania opcji ($S > K$), czyli opcja jest *in the money* i możliwe jest jej wykonanie.

$$P(S, t) = N\left(\frac{-\ln \frac{B}{S} + \left(r - q - \frac{1}{2}\sigma^2\right)(T - t)}{\sigma\sqrt{T - t}}\right) - \left(\frac{K}{S}\right)^{\frac{2(r - q - \frac{1}{2}\sigma^2)}{\sigma^2}} N\left(\frac{-\ln \frac{B}{S} + 2\ln \frac{K}{S} + \left(r - q - \frac{1}{2}\sigma^2\right)(T - t)}{\sigma\sqrt{T - t}}\right), \quad (1.10)$$

gdzie:

B – oznacza górną barierę, która musi zostać przekroczona, aby wykonanie opcji było możliwe (bądź jeżeli taka bariera nie została w programie ustalona, oznacza wysokość ceny akcji jaka była osiągana w momencie wykonania opcji przez pracowników o podobnej pozycji w hierarchii przedsiębiorstwa w przeszłości), wartość ta ustalana jest na podstawie danych historycznych.

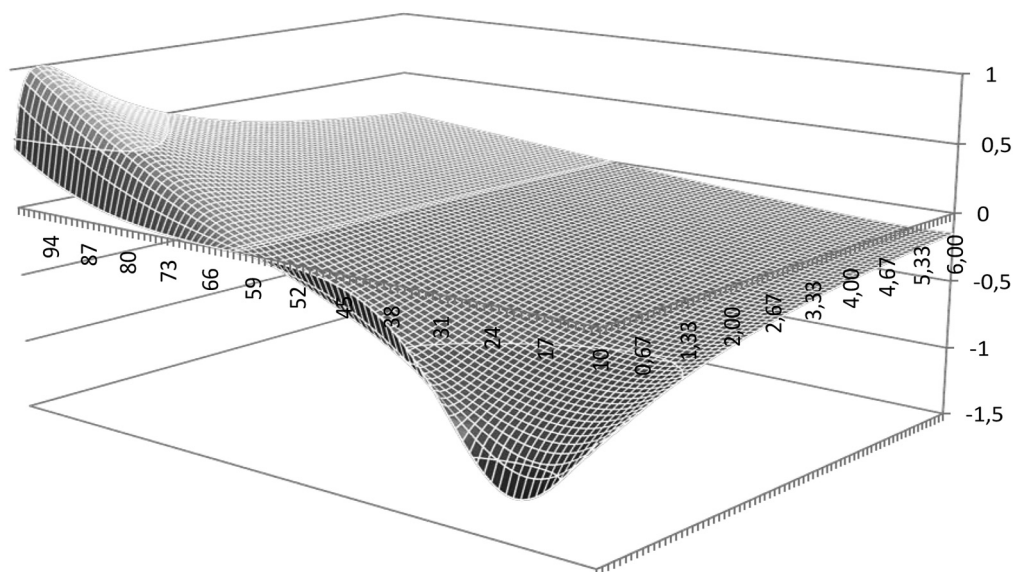
Jednak w celu oszacowania wskaźnika odejść, w badaniach wykorzystywane zostanie pewne przekształcenie tej funkcji. Wzór (1.10) $P(S, t)$ jest funkcją określającą stopę wczesnego wykonania opcji, natomiast w moich badaniach będzie potrzebne oszacowanie funkcji $\lambda(S, t)$. W tym celu wykorzystany zostanie wzór (1.11) [21].

$$e^{-\lambda(S,t)(T-t)} = 1 - P(S,t) \quad (1.11)$$

W celu oszacowania funkcji $\lambda(S, t)$ potrzebne jest przekształcenie wzoru (1.11). Wzór po przekształceniu przedstawiony jest jako wzór (1.12).

$$\lambda(S,t)(T-t) = \ln\left(\frac{1}{1-P(S,t)}\right) \quad (1.12)$$

Tak opracowana funkcja $\lambda(S, t)$ wykorzystana zostanie w badaniach. Przy tworzeniu funkcji prawdopodobieństwa użyto parametru B , który pozwolił na włączenie do rozważań bariery ograniczającej możliwość wykonania opcji. Potrzebne jest jednak wprowadzenie do modelu wskaźnika odejść, obliczonego na podstawie funkcji prawdopodobieństwa zaproponowanej w publikacji Musieli i Rutkowskiego [13]. Ponieważ funkcja $\lambda(S,t)$ przyjmuje duże wartości, dlatego zdecydowano się na normalizację tej zmiennej względem maksimum, poprzez podzielenie wszystkich otrzymanych wartości przez maksymalną otrzymaną wartość tego wskaźnika. Wykres funkcji $\lambda(S, t)$ opracowany dla danych z badanego przeze mnie programu dla spółki ComArch S.A. przedstawiony został na wykresie 1.3.



Rysunek 1.3 Znormalizowane wartości funkcji $\lambda(S, t)$

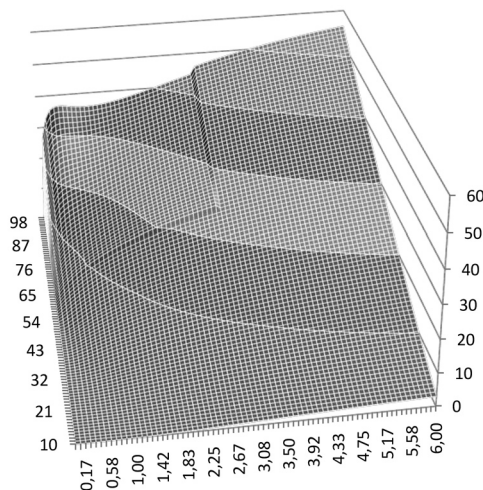
6. PRZYKŁAD

Praktyczne wykorzystanie zaprezentowanego modelu wyceny opcji menedżerskich sporządzone zostało dla akcji spółki ComArch S.A. Skonstruowany program menedżer-

ski trwa od 11.02.2004 roku do 11.02.2010 roku. Okres nabywania uprawnień wynosi 3 lata. Wskaźnik wielokrotności z uwagi na stosunkowo wysoką zmienność ceny akcji przy cenie wykonania opcji (62,5 zł) oraz cenie akcji w 2004 roku (46,8 zł) ustalono na poziomie 2,2 ceny wykonania opcji zgodnie z badaniami Huddarta i Langa [7], gdyż program z założenia obejmuje pracowników, a nie tylko najwyższych rangą menedżerów. Cenę wykonania ustalono na podstawie obserwacji kształtowania się ceny akcji spółki ComArch S.A. w okresie od 10.03.1999 roku do 11.02.2004 roku. Stopa wolna od ryzyka na podstawie oprocentowania 52-tygodniowych bonów skarbowych wynosi w prezentowanych modelach 3,85%. Stopę odejścia w okresie otwartym ustalono na poziomie 5%. Jako stopę odejścia w okresie otwartym wykorzystano funkcyjną postać wskaźnika odejść (która jest w poszczególnych punktach przekształcana zgodnie ze wzorem 1.1). Najważniejsze parametry obliczone w przypadku konstruowania drzewa trójmianowego przyjęły następujące wartości: $p_u = 0,210664$, $p_d = 0,292701$, $p_m = 0,496635$.

Na podstawie modelu drzewa trójmianowego i przedstawionych danych wejściowych obliczono cenę opcji w momencie przyznania na poziomie 22,1877 zł (przy cenie akcji spółki ComArch S.A. wynoszącej 11.02.2004 roku 46,8 zł). Wykorzystując jawną metodę różnic skończonych (stworzoną na podstawie artykułu Westa [21]) dla dokładnie takich samych danych wejściowych oraz warunków programu, otrzymana wartość jest równa tej z modelu drzewa trójmianowego. Stanowi to o poprawności i dużej dokładności zaprezentowanego modelu.

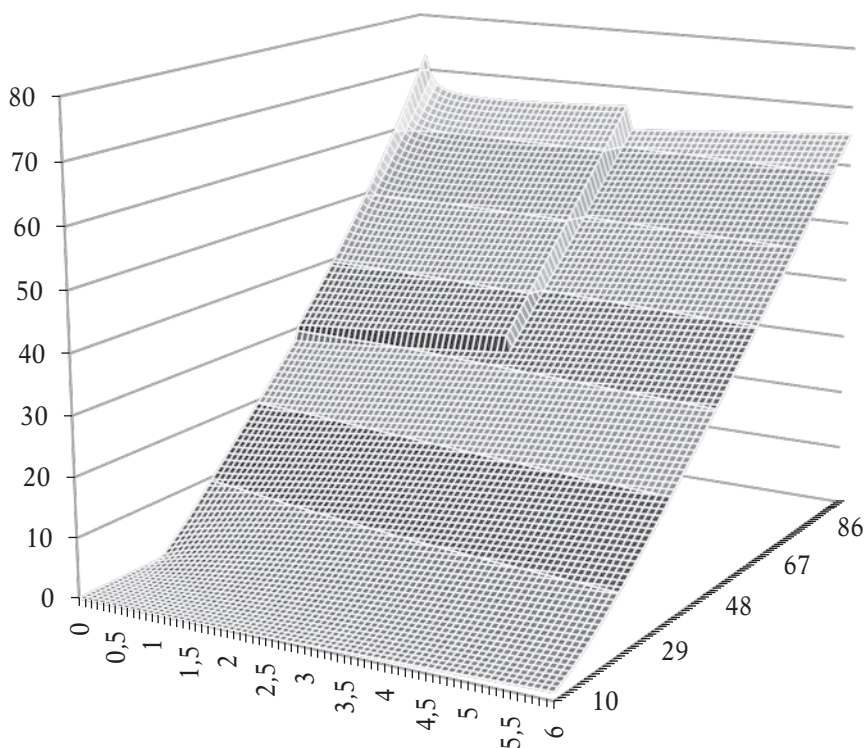
Na wykresie 1.4 przedstawione zostały badania symulacyjne ceny opcji menedżerskiej dla spółki ComArch S.A. Oś czasu przedstawia liczbę lat do wygaśnięcia opcji. Charakterystyczne załamanie powierzchni spowodowane jest uwzględnieniem wskaźnika odejść w okresie otwartym (między 3 a 6 rokiem trwania programu motywacyjnego) dla ceny akcji umożliwiającej jej wykonanie w związku z ustalonym wskaźnikiem wielokrotności (2,2).



Rysunek 1.4 Symulacja wartości opcji dla spółki ComArch S.A.

Drugi przykład obliczeniowy przedstawia stworzony dla spółki PKN Orlen S.A. program motywacyjny przeznaczony dla najwyższych rangą menedżerów. Program obejmuje okres od 29.11.2002 roku do 28.11.2008 roku. Okres nabywania uprawnień ustalony został jako 3 lata, natomiast wskaźnik wielokrotności, zgodnie z badaniami Carpenter wynosi 2,8 (z uwagi na fakt objęcia programem tylko najwyższych rangą menedżerów). Cena akcji w momencie przyznania wynosi 17,15 zł, natomiast cena wykonania – 25 zł. Parametry wykorzystywane przy tworzeniu drzewa trójkątnego przyjęły wartości: $p_u = 0,239137$, $p_d = 0,261104$, $p_m = 0,499759$.

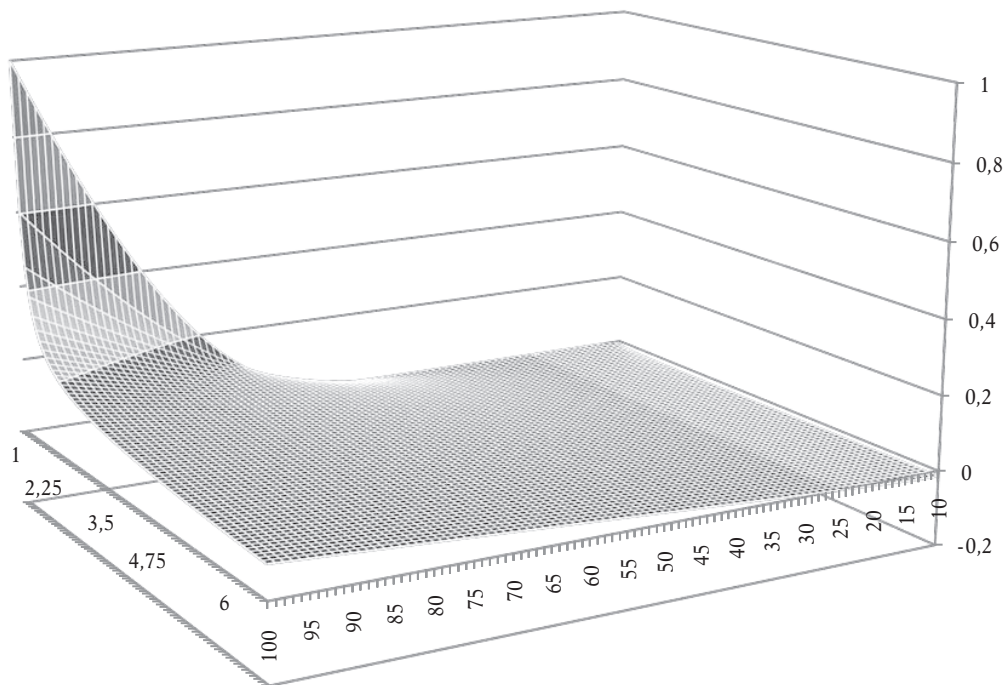
Wykorzystując zaprezentowany model drzewa trójkątnego dla skonstruowanego na podstawie powyższych danych programu menedżerskiego dla spółki PKN Orlen S.A. obliczono cenę opcji w momencie przyznania na poziomie 4,0122 zł (przy cenie akcji w tym momencie wynoszącej 17,15 zł). Rysunek 1.5 przedstawia znormalizowane wartości funkcji $\lambda(S, t)$, czyli postaci funkcyjnej wskaźnika odejść dla analizowanego programu spółki PKN Orlen S.A.



Rysunek 1.5 Znormalizowane wartości funkcji $\lambda(S, t)$ dla programu spółki PKN Orlen S.A.

Badania symulacyjne wartości opcji dla programu menedżerskiego dla spółki PKN Orlen S.A. przedstawione zostały na rysunku 1.6. W przypadku tego programu widoczne załamanie wartości opcji spowodowane jest również uwzględnieniem w okresie otwar-

tym dla odpowiedniej wartości akcji wskaźnika odejść. W przypadku przedstawionego na rysunku 1.6 wykresu, można zauważyć, wyraźniej niż w przypadku programu dla spółki ComArch S.A., zwiększenie wartości opcji już na początku okresu otwartego, co stanowi dla menedżera premię za zwiększone ryzyko niewykonania opcji.



Rysunek 1.6 Symulacja wartości opcji dla spółki PKN Orlen S.A.

7. PODSUMOWANIE

Z programami menedżerskimi związanych jest wiele problemów. Pierwszą grupą problemów stanowią zagadnienia związane z odpowiednią konstrukcją samych programów. Z jednej strony program menedżerski ma spełniać oczekiwania wystawiającego je przedsiębiorstwa, z drugiej jednak ma stanowić wynagrodzenie dla menedżera za osiągnięte wyniki. Najważniejszym zadaniem na etapie konstruowania programu jest zatem takie zdefiniowanie warunków, aby możliwe było ich spełnienie, ale równocześnie taka konstrukcja samego programu, aby niemożliwe było omówione wcześniej nieuczciwe wpływanie zainteresowanych osób na kształtowanie zysku z opcji menedżerskich.

Drugą grupę problemów związanych z programami menedżerskimi stanowi obszar związany z ich odpowiednią wyceną. W przedstawionym modelu starano się pokazać jak uwzględnić konkretne warunki rynkowe w procesie wyceny. Oczywiście możliwości związanych ze swobodnym kształtowaniem warunków jest tyle, że niemożliwym staje

się przedstawienie w jednym artykule wszystkich warunków, jednak postępując analogicznie do zaprezentowanych rozwiązań, z pewnością można poradzić sobie z bardzo egzotycznymi wymaganiami postawionymi menedżerom. Kolejnym problemem związanym z odpowiednią wyceną programów menedżerskich jest dostosowanie modeli do wymagań Międzynarodowego Standardu Sprawozdawczości Finansowej 2. Zrozumienie standardu stanowi absolutną podstawę przy konstruowaniu modeli wyceny. Jednak zakres niniejszego artykułu nie pozwala na szersze omówienie kwestii poruszonych w MSSF 2 ze względu na szczegółowość całego standardu i związaną z tym obszerność. Przedstawiony model jest zmodyfikowany pod kątem tegoż standardu.

Niniejszy artykuł miał na celu przekrojowe przedstawienie zagadnienia programów opcji menedżerskich od momentu konstrukcji programu do jego wyceny. Każdy z punktów mógłby zostać znacznie obszerniej potraktowany, jednak brak zbiorczych opracowań dotyczących opcji menedżerskich w polskiej literaturze skłania do przedstawienia przekroju całej tematyki bardziej niż szczegółowego zajęcia się wycinkiem tej problematyki. Przedstawiony model wyceny jest bardzo elastyczny w odniesieniu do uwzględnianych w programie warunków, a konkretne formy modelu zaprezentowane w tym artykule, miały posłużyć przykładowemu przedstawieniu możliwości modelu.

Uniwersytet Ekonomiczny w Poznaniu

LITERATURA

- [1] Ayache E., Forsyth P.A., Vetzal K.R., [grudzień 2002], *Next generation models for convertible bonds with credit risk*, Wilmott Magazine, s. 68-77.
- [2] Ayache E., Forsyth P.A., Vetzal K.R., [2003], *Valuation of convertible bonds with credit risk*, „Journal of Derivatives”, jesień 2002, s. 9-30.
- [3] Carpenter J., [1998], *The exercise and valuation of executive stock options*, „Journal of Financial Economics”, nr 48, s. 127-158.
- [4] Carr P., Linetsky V., [2000], *The Valuation of Executive Stock Options in an Intensity-Based Framework*, „European Finance Review”, nr 4, s. 211-230.
- [5] Derman E., Kani I., Chriss N., [luty 1996], *Implied Trinomial Trees of the Volatility Smile*, Goldman Sachs Quantitative Strategies Research Notes.
- [6] Heron R.A., Lie E., [2006], *Does backdating explain the stock price pattern around executive stock option grants?*, „Journal of Financial Economics”, s. 3.
- [7] Huddart S., Lang M., [1996], *Employee Stock Option Exercises: An Empirical Analysis*, „Journal of Accounting and Economics”, Vol. 21, nr 1, s. 5-43.
- [8] Hull J., White A., [2002], *How to value Employee Stock Options*.
- [9] Hull J., White A., [2003], *Accounting for Employee Stock Options*.
- [10] Kuźmierkiewicz M., [1999], *Ewolucja rynku opcji ku pozagiędkowym opcjom egzotycznym i ich klasyfikacja*, „Bank i Kredyt” nr 3, s. 18.
- [11] Lakonishoka J., Lee I., [wiosna 2001], *Are Insider Trades Informative?*, „Review of Financial Studies”, Vol. 14, nr 1.
- [12] Międzynarodowy Standard Sprawozdawczości Finansowej 2 (MSSF 2, IFRS 2), 19.02.2004 r.
- [13] Musiela M., Rutkowski M., [Springer 2005], *Martingale Methods in Financial Modelling, Second Edition*.
- [14] Narayanan M.P., Seyhun H.N., [2007], *Dating Games: Do Managers Designate Grant Dates to Increase their Compensation*, Review of Financial Studies.
- [15] Noreen E., Wolfson M., [1981], *Equilibrium warrant pricing models and accounting for executive stock options*, „Journal of Accounting Research”, nr 29, s. 384-398.

- [16] Ong M., [1996], *Exotic options: The market and their taxonomy* (rozdział I), *The handbook of exotic options: instruments, analysis, and applications*, red. Nelken Israel, McGraw-Hill Book Company, New York 1996, s. 3, 10-11.
- [17] Preuss P., Ślusarski W., [1 marca 2007], *Ujęcie księgowe i wycena programów menedżerskich*.
- [18] Rubinstein M., [1995], *On the accounting valuation of employee stock options*, „Journal of Derivatives”, nr 3, s. 8-24.
- [19] Seyhuna H.N., [1988], *The Information Content of Aggregate Insider Trading*, „Journal of Business”, nr 61, s. 1-24.
- [20] Seyhuna H.N., [1992], *Why Does Aggregate Insider Trading Predict Future Stock Returns*, „Quarterly Journal of Economics”, Vol. 107, nr 4, s. 1303-1331.
- [21] West G., [14 stycznia 2009], *A Finite Difference Model for Valuation of Employee Stock Options*.

Praca wpłynęła do redakcji w lipcu 2010 r.

ZASTOSOWANIE MODELU DRZEWA TRÓJMIANOWEGO DO WYCENY OPCJI MENEDŻERSKICH

Streszczenie

W artykule omówiono konstrukcję oraz wycenę programów opcji menedżerskich. Przedstawione zostały warunki rynkowe oraz nierynkowe wykorzystywane w przypadku konstruowania programów opcyjnych. Wykorzystana metoda wyceny została oparta na drzewie trójmianowym, które zostało zmodyfikowane umożliwiając włączenie do wyceny zaprezentowanych warunków nabycia uprawnień oraz dostosowując model wyceny do wymagań Międzynarodowego Standardu Sprawozdawczości Finansowej 2.

Słowa kluczowe: opcje menedżerskie, wskaźnik odejść, wskaźnik wielokrotności

PRICING OF EMPLOYEE STOCK OPTIONS USING TRINOMIAL TREE

Summary

The article discusses the construction and pricing of Employee Stock Options Plans. There were presented performance conditions and market conditions that should be taken into account constructing ESOPs. The presented valuation method is based on a trinomial tree, which was modified making it possible to include presented vesting conditions in the pricing model and adapting it to the requirements of International Financial Reporting Standard 2.

Key words: Employee Stock Options, employee exit rate, multiple rate

BARBARA BĘDOWSKA-SÓJKA, AGATA KLIBER¹

REALIZED VOLATILITY VERSUS GARCH AND STOCHASTIC VOLATILITY MODELS. THE EVIDENCE FROM THE WIG20 INDEX AND THE EUR/PLN FOREIGN EXCHANGE MARKET

1. INTRODUCTION

Volatility is an unobservable feature of each financial market. While we can directly observe the prices of instruments and their movement, we cannot observe their volatility. Volatility is a measure associated with risk and uncertainty connected with unexpected changes of an instrument price. This makes financial time series unpredictable. Since we cannot observe volatility, we only approximate it with the help of statistical models. The simplest and most naive measure of volatility of a given instrument is the standard deviation (or variance) of its price. However, the measure is static while the volatility of an instrument changes on day-to-day basis. That is why dynamic measures of volatility have gained more and more popularity.

Dynamic modelling of volatility can be divided into three groups [9]:

- taking advantage of volatility models such as the (Generalised) Autoregressive Conditional Heteroskedasticity models ((G)ARCH) and the Stochastic Volatility models (SV),
- implied volatility models, deriving volatility from the market option prices,
- realized volatility (RV), based upon the high-frequency data.

The aim of the article is to compare the estimates of volatility obtained from the parametric volatility models, namely the GARCH and the SV, with realized volatility RV as a non-parametric measure. We consider RV measures obtained from data of different frequencies and calculated within two approaches: with and without night returns. We take into account the Polish stock exchange market and the foreign exchange one: our sample consist of the WIG20 index and the EUR/PLN exchange rate and covers a hectic crisis period.

The remainder of the paper is organized as follows. Section 2 describes GARCH and SV models. The different types of calculating realized volatility are presented in Section 3. Section 4 presents the data and empirical results of estimation of the GARCH and the SV models. The comparison of realized volatility measures and in-sample estimated conditional volatility are shown in Section 5. Section 6 presents

¹ The work is supported by MNiS through the Project NN 111 1256 33. We are very thankful to anonymous referee for helpful suggestions and remarks. All remaining errors are ours.

accuracy of forecasts obtained from the models in comparison to realized volatility. Section 7 concludes.

2. PARAMETRIC ESTIMATES OF VOLATILITY: GARCH AND SV MODELS

Although the financial time series are unpredictable, they display both temporal dependency in their second order moments and heavy-picked and tailed distributions. This phenomenon has been known since the work of Mandelbrot [17], but only after the introduction of the autoregressive conditionally heteroskedastic models the econometricians started modelling the dynamics of volatility.

The first model from the group of autoregressive models was the Autoregressive Conditional Heteroskedasticity (ARCH) model, proposed by Engle [11]. Let us define:

$$r_t = 100 \cdot \ln(S_t/S_{t-1}), \quad (1)$$

where r_t denotes the logarithmic return, and S_t the price of the instrument at time t .

Next, let:

$$y_t = r_t - E(r_t | \mathcal{F}_{t-1}), \quad (2)$$

where $E(r_t | \mathcal{F}_{t-1})$ denotes the expected conditional value of the instrument price (usually modelled with ARMA models). The assumption of the ARCH(q) model is that the time series $\{y_t\}$ is not correlated, but the variables are dependent and this dependence can be described with the help of a function whose arguments are the squared values of the lagged variables y_t :

$$y_t = \sigma_t \varepsilon_t, \\ \sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \dots + \alpha_q y_{t-q}^2.$$

Where: $\{\varepsilon_t\} \sim iid(0,1)$ – independent random variables of the same distribution, $\alpha_0 > 0$, $\alpha_i \geq 0$, for $i > 0$ and σ_t^2 denotes the unobserved volatility (i.e. conditional variance).

A generalisation of the ARCH model was proposed by Bollerslev [4]. Again, it is assumed that the average of the logarithmic returns can be described with the help of the ARMA model. The GARCH(p, q) model takes the following form:

$$y_t = \sigma_t \varepsilon_t, \\ \sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2,$$

where: $\{\varepsilon_t\} \sim iid(0,1)$, $\alpha_i \geq 0$, $\beta_j \geq 0$, $i = 1, \dots, p$, $j = 1, \dots, q$.

The advantage of the GARCH model over the ARCH is that the latter is often overparametrized, whereas in the practice the most common is plain vanilla GARCH(1,1).

Since the first GARCH(p, q) model was introduced, there have emerged lots of its modifications, which take into account many additional features of time series (such as leverage effect, long memory, different error distributions, see in: [5], [23]).

In many financial series the conditional variance estimated through GARCH(p, q) reveal strong persistence, characterized by:

$$\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j = 1.$$

In this case we estimate IGARCH model, which can be written as:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i}^2 + \sum_{j=1}^q (1 - \alpha_i) \sigma_{t-j}^2.$$

Under the IGARCH(1,1) the unconditional variance of y_t is not defined.

2.1. GARCH MODELS WITH NORMAL AND STUDENT DISTRIBUTIONS

For the standard normal distribution, and if we express the conditional mean equation as in Equation (2) and $y_t = \sigma_t \varepsilon_t$, the log-likelihood function is given by:

$$L_{norm} = -\frac{1}{2} \sum_{t=1}^T \left[\ln(2\pi) + \ln(\sigma_t^2) + \varepsilon_t^2 \right]$$

with T denoting the number of observation.

Due to the existence of fat tails in financial time-series, the use of a Student distribution in combination with GARCH models is suggested. In case of a Student distribution, the log-likelihood function is given by:

$$L_{Student} = T \left\{ \ln \Gamma\left(\frac{\nu+1}{2}\right) - \ln \Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2} \ln[\pi(\nu-2)] \right\} \\ - \frac{1}{2} \sum_{t=1}^T \left[\ln(\sigma_t^2) + (1+\nu) \ln\left(1 + \frac{\varepsilon_t^2}{\nu-2}\right) \right],$$

where ν denotes the degrees of freedom ($2 < \nu < \infty$) is estimated together with other parameters in the model and $\Gamma(\cdot)$ is the gamma function.

2.2. GARCH MODELS WITH LEVERAGE EFFECT

This effect was first described by Black [3] and it takes into account the fact that high negative returns are accompanied by higher volatility than the positive returns of the same magnitude. A volatility model commonly used in the case of leverage effect is the threshold GARCH, presented in Glosten, Jagannathan and Runkle [12]. The conditional variance equation takes the following form:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p (\alpha_i + \gamma_i N_{t-i}) y_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2,$$

where N_{t-i} is a dummy variable indicating the sign of innovations:

$$N_{t-i} = \begin{cases} 1 & \text{if } y_{t-i} < 0 \\ 0 & \text{if } y_{t-i} \geq 0 \end{cases}.$$

The parameter γ is interpreted as a leverage effect and it counts for any asymmetry on the volatility response to negative and positive shocks. If γ is positive, the past negative shocks have a larger impact on conditional variance than the positive shocks.

2.3. UNIVARIATE STOCHASTIC VOLATILITY MODELS

In the ARCH-type models volatility is made dependent on the variability of the past observations. It means, that volatility at the moment t is determined by the variables, the values of which are known up to time $t - 1$. An alternative approach was proposed by Clark [7] and Taylor [25]. In Stochastic Volatility models volatility is driven by additional, unobserved components (see e.g. [18]).

2.3.1. THE BASIC STOCHASTIC VOLATILITY MODEL

Let h_t^2 denote the latent volatility on the day t , while $\phi \in (-1; 1)$ be a correlation coefficient. The basic SV model can be written in the following form:

$$\begin{aligned} y_t &= \exp\left(\frac{1}{2}h_t\right)u_t, u_t \sim iid N(0, 1) \\ h_t &= \alpha + \phi(h_{t-1} - \alpha) + v_t, v_t \sim iid N(0, \tau^2), \end{aligned}$$

where $h_0 \sim N(\alpha, \tau^2)$.

The parameter ϕ is usually interpreted as so called volatility persistence parameter, while τ^2 is interpreted as *volatility of log-volatility* (see e.g. [15], [27]).

2.3.2. STOCHASTIC VOLATILITY MODEL WITH THE STUDENT DISTRIBUTION

The basic model can be changed into more sophisticated ones. One can incorporate into it the knowledge about the data characteristics, e.g. non-normality of distribution. In the SV model with the Student distribution it is assumed that the price of the instrument follows the Student distribution with ν degrees of freedom:

$$y_t = \exp\left(\frac{1}{2}h_t\right)u_t, u_t \sim iid t_\nu,$$

$$h_t = \alpha + \phi(h_{t-1} - \alpha) + v_t, v_t \sim iid N(0, \tau^2).$$

2.3.3. STOCHASTIC VOLATILITY MODEL WITH LEVERAGE EFFECT

The model can be made even more sophisticated if one wishes to include the leverage effect into it.

The model takes the following form:

$$y_t = \exp\left(\frac{1}{2}h_t\right)u_t, t = 1, \dots, N,$$

$$h_{t+1} = \alpha + \phi(h_t - \alpha) + \tau v_{t+1}, t = 1, \dots, N-1,$$

$$\begin{pmatrix} u_t \\ v_{t+1} \end{pmatrix} \sim iid. N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right].$$

Then, (h_t) and returns (y_t) have the following distributions:

$$h_{t+1} | h_t, \alpha, \phi, \tau^2 \sim N(\alpha + \phi(h_t - \alpha), \tau^2),$$

$$y_t | h_{t+1}, h_t, \alpha, \phi, \tau^2, \rho \sim N\left(\frac{\rho}{\tau} e^{h_t/2} (h_{t+1} - \alpha - \phi(h_t - \alpha)), e^{h_t} (1 - \rho^2)\right).$$

If $\rho < 0$, this means that the leverage effect exists and volatility tends to increase more when large negative returns are present in the market. The parameter ϕ denotes so called persistence of the volatility, which is said to be high if the absolute value of this parameter is close to 1.

The model presented above is the one proposed in [13] and is an alternative to the model presented in [14]. The comparison of the models can be found in [26]. For more examples of univariate Stochastic Volatility Models and their implementations in WinBugs we refer the Reader to [27].

3. REALIZED VOLATILITY

Realized volatility is a non-parametric measure proposed by Andersen and Bollerslev [2] as a proxy for non-observable integrated volatility. Let us introduce the basic concept of integrated volatility (the whole section is based mainly on [16]).

We consider the following diffusion process for the logarithmic price $\ln S_t$:

$$d \ln S(t) = \alpha(t) dt + \sigma(t) dW(t), \quad t \geq 0,$$

where, as previously:

$S(t)$ – asset price at time t ,

$d \ln S(t)$ – continuously compounded return,

$W(t)$ – standard Brownian motion process,

$\mu(t)$ – drift,

$\sigma(t)$ – volatility.

For very small time intervals, Δ , we obtain:

$$r(t, \Delta) \equiv S(t) - S(t - \Delta) \approx \alpha(t - \Delta) \Delta + \sigma(t - \Delta) \Delta W(t),$$

where $\Delta W(t)$ is normally distributed with zero mean and standard deviation equal to Δ .

When daily one-period return is considered:

$$r_t \equiv S(t) - S(t - 1) = \int_{t-1}^t \alpha(s) ds + \int_{t-1}^t \sigma(s) dW(s).$$

Additionally, conditionally on the sample path of the drift and the spot volatility,

$$r_t \sim N\left(\int_{t-1}^t \alpha(s) ds, \int_{t-1}^t \sigma^2(s) ds\right),$$

where $\int_{t-1}^t \sigma^2(s) ds$ is so-called integrated volatility (integrated variance).

However, integrated volatility is not observed in practice. Realized volatility provides a consistent non-parametric estimate of a financial instrument prices variability over a given time interval and therefore is often treated as a proxy for integrated volatility.

The realized volatility is denoted $RV_t(\Delta)$ and defined as:

$$RV_t(\Delta) \equiv \sum_{i=1}^N r_{t,i}^2,$$

where:

Δ – is a number of equally spaced intervals within a day,

$r_{t,i}$ – is a logarithmic return on day t at time interval i ; with $i = 1, \dots, N$ and $t = 1, \dots, T$.

In the present work realized volatility is calculated in two ways both with and without night return (RV1 and RV2 respectively).

As a non-parametric and easy to obtain estimate, realized volatility is very commonly used in assessing the quality of conditional variance forecasts from the GARCH and the SV models within Mincer-Zarnowitz regression [19], which is in the form of ex-post volatility regression:

$$\check{\sigma}_t^2 = \alpha_0 + \alpha_1 \hat{\sigma}_t^2 + a_t,$$

where:

$\check{\sigma}_t^2$ is ex-post volatility (e.g. realized volatility) at time t ,

$\hat{\sigma}_t^2$ is estimated (in-sample) or forecasted (out-of sample) volatility at time t ,

a_t – independent and identically distributed; $a_i \sim N(0,1)$.

α_0 and α_1 are parameters to be estimated.

If the model for conditional variance is well specified, and if $E(\check{\sigma}_t^2) = \sigma_t^2$, we should have: $\alpha_0 = 0$, $\alpha_1 = 1$ [2]. According to specific features of financial data series the value of R^2 is usually low (even less than 5%).

Our approach is to use realized volatility for two purposes. First, we check if realized volatility is close to in-sample parametric estimates of volatility from GARCH and SV models. Having in-sample estimates we are able to compare two different measures of realized volatility, one of which includes night return and the other does not. In this case $\check{\sigma}_t^2$ is the realized volatility and $\hat{\sigma}_t^2$ is the estimated conditional variance from the GARCH or the SV models.

Second, we use GARCH and SV models forecasts to examine which realized volatility, including night return or not, from lower or higher frequencies, is closer to forecasts obtained from these models. In this case $\hat{\sigma}_t^2$ stands for the forecast of conditional variance from the GARCH or the SV models.

4. DATA DESCRIPTION AND EMPIRICAL RESULTS

4.1. DATA

Our data set consists of daily and intradaily prices of the WIG20 index and the EUR/PLN exchange rate over the period 2006-01-02 to 2009-11-06. We transform them into percentage logarithmic returns. The data are taken from www.stooq.pl. We estimate conditional variance from both the GARCH model (starting at 2006-01-10) and the SV model (starting at 2008-01-03) and introduce them into the Mincer-Zarnowitz in-sample regression together with the realized volatility from 2008-01-02 to 2009-07-31. Finally, Mincer-Zarnowitz regression is performed on the forecasts of the conditional variance from the GARCH and the SV models – there are 70 forecast from 2009-08-01 to 2009-11-06.

All calculations were done using OxMetrics 6.0 [10] and WinBugs 1.4. In case of WinBugs we took advantage of the models written by Yu [26, 27] and available at his webpage.

4.2. ESTIMATION OF GARCH MODELS

The GARCH model was chosen by taking into account information criteria. Four models were compared: the most popular GARCH(1,1) with the normal and the Student distributions and the GJR-GARCH model [12] with normal and Student distributions. Here we present values of the Schwarz Information Criterion. Other considered criteria (Akaike and Hannan-Quinn) give the same indication.

Table 1

Comparison of GARCH models estimated for the WIG20 index

Model	GARCH(1,1)	GARCH(1,1)	GJR-GARCH(1,1)	GJR-GARCH(1,1)
distribution	normal	Student- <i>t</i>	normal	Student- <i>t</i>
SIC	4,075	4,071	4,067	4,065

The GJR model with the Student distribution occurred to be the best. In Table 2 the parameters estimated for GARCH(1,1) and GJR-GARCH(1,1) (both with the Student distribution) are presented.

Table 2

Estimates for GARCH(1,1) and GJR-GARCH(1,1) for returns of the WIG20 index

	α_0	α_1	β_1	γ	v	LL
GARCH(1,1)	0,050	0,054	0,933		11,037	-1789,23
	(0,025)	(0,014)	(0,016)			
GJR-GARCH(1,1)	0,051	0,005	0,945	0,071	12,316	-1796,60
	(0,026)	(0,011)	(0,019)	(0,023)		

For the GARCH(1,1) we obtain statistically significant estimates of the parameters. In the GJR-GARCH(1,1) model we obtain statistically significant parameter of the leverage effect and statistically insignificant parameter of the past squared shocks. Introducing leverage effect into conditional variance equation also resulted in increasing the degrees of freedom in the Student distribution.

Table 3

Comparison of GARCH models estimated for the EUR/PLN exchange rate

Model	GARCH(1,1)	GARCH(1,1)	IGARCH(1,1)	IGARCH(1,1)
distribution	normal	Student- <i>t</i>	normal	Student- <i>t</i>
SIC	1,639	1,622	1,632	1,615

In the case of the EUR/PLN exchange rate the IGARCH model with the Student distribution occurred to be the best. The estimates are presented in Table 4 together with the standard GARCH(1,1) model.

Table 4

Estimates of GARCH(1,1) and IGARCH(1,1) for the EUR/PLN exchange rate

	α_0	α_1	β_1	v	LL
GARCH(1,1)	0,02	0,08	0,92	7,75	-709,67
	(0,01)	(0,02)	(0,02)		
IGARCH(1,1)	0,02	0,08	0,92	7,51	-709,73
	(0,01)	(0,02)			

4.3 ESTIMATION OF STOCHASTIC VOLATILITY MODELS

As an alternative to GARCH we estimated the Stochastic Volatility model. We compared three models: the basic Stochastic Volatility model, the model with leverage effect and the model with the Student distribution.

Since in case of SV models, distribution of the time series at moment t is unknown conditional upon the observations up to the moment $t - 1$, estimation of the models' parameters with the help of the standard methods based upon the maximum-likelihood is not possible. Broto and Ruiz [6] provide a survey of methods of the SV models estimations, while Kim et al. analyze the efficiency of their estimation [15]. From the various methods mentioned in the papers we chose Bayesian method based upon the MCMC (Markov Chain Monte Carlo) algorithm [27]. An example of the application of the Bayesian methodology to the estimation of stochastic volatility processes in analysis of the financial time series can be found e.g. in: [8], [21] and [22].

For each of the series we estimated three mentioned models and compared them with the help of DIC (Deviance Information Criterion). The criterion was first introduced by Spiegelhalter [24] as a measure of model comparison and adequacy. It is given by the following expression:

$$DIC(m) = 2\overline{D(\theta_m, m)} - D(\overline{\theta_m}, m) = 2D(\overline{\theta_m}, m) + 2p_m,$$

where:

$$D(\bar{\theta}_m, m) = -2 \log f(y | \theta_m, m)$$

and $\overline{D(\theta_m, m)}$ is the posterior mean of the deviance measure $D(\bar{\theta}_m, m)$. The value p_m can be interpreted as the number of “effective” parameters of the model m : $p_m = \overline{D(\theta_m, m)} - D(\bar{\theta}_m, m)$. Eventually, $\bar{\theta}_m$ is the posterior mean of parameters involved in the model m [20]. The criterion is considered as a generalization of the Akaike Information Criterion (AIC) [1]. The minimum value of the criterion indicates the model which offers the best short-term predictions. Since we are interested especially in the prediction power of the models, we report here the values of DIC and use the criterion to select the best model. For the Reader’s information we report also the values of the Bayes Information Criterion (BIC) which is based upon the Schwarz criterion [9]. The BIC for the model m is defined as:

$$BIC(m) = D(\hat{\theta}_m, m) + d_m \log(n),$$

where $\hat{\theta}_m$ are the maximum likelihood estimates of parameters θ_m of model m and d_m is the dimension of θ_m . As previously, $D(\hat{\theta}_m, m)$ is the deviance measure of model m .

The tables below present the results of models comparison:

Table 5

Comparison of SV models estimated for the WIG20 index

Model	basic SV	t -SV	leverage SV
BIC	1768.108	1777.850	1465.22
DIC	1756.880	1760.980	1678.030

Table 6

Comparison of SV models estimated for the EUR/PLN exchange rate

Model	basic SV	t -SV	leverage SV
BIC	916.610	959.512	755.830
DIC	913.312	982.680	860.019

What is interesting, in both cases the basic SV model performed better than the model with the Student distribution. DIC criterion reached its minimum for the models with the leverage effect. Thus, we further present the results only from this model.

Since the estimation was performed using the WinBUGS software for Bayesian modeling, we treat each parameter of the model as a random variable and present

the obtained mean, median, standard deviation and 95% confidence interval for each of them. The MC error reported in each table denotes the so called Monte Carlo error which measures the variability of each estimate due to the simulation (see e.g.: [20]). We also present the obtained posterior densities of each parameter. In the case of stochastic volatility we present the 95% confidence interval of the estimated values.

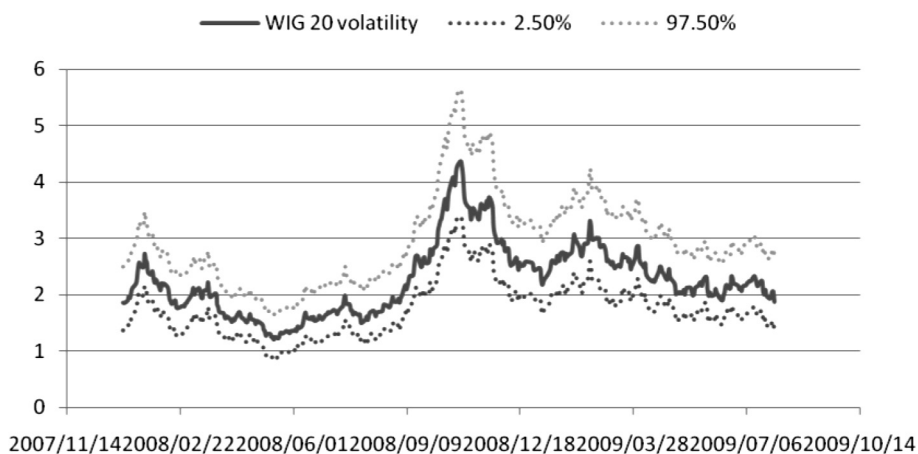


Figure 1. 95% confidence intervals obtained for volatility of WIG20 (leverage SV model)

Table 7

Descriptive statistics and 95% confidence interval for parameters of the WIG20 index volatility

Parameter	Mean	Standard deviation	MC error	2.5%	Median	97.5%
τ	0.132	0.025	0.002	0.089	0.129	0.196
ρ	-0.582	0.154	0.012	-0.814	-0.606	-0.193
ϕ	0.982	0.010	0.001	0.956	0.983	0.996
α	0.024	0.016	0.001	0.003	0.021	0.065
α	1.252	0.301	0.020	0.612	1.272	1.806

To estimate the parameters of the model we took advantage of the programs written by Yu [26, 27]. In case of the three SV models the parameters: τ , ϕ and α have the following prior distributions: τ – the inverse Gamma, α – the Normal, while $\phi^* = (\phi + 1)/2$ – the Beta distribution. The additional parameter ρ , present in the leverage SV model is uniformly distributed: $\rho \sim U(-1, 1)$.

Based upon the results displayed in Table 7 we can state that the leverage effect is present in the volatility of the WIG20 index. The parameter ρ is negative and both ends of the 95% confidence interval are negative. At the same time we can observe that the value of the parameter ϕ is close to 1, which is a sign of a high persistence of volatility. Figures 2 to 6 present the obtained posterior densities of each parameter.

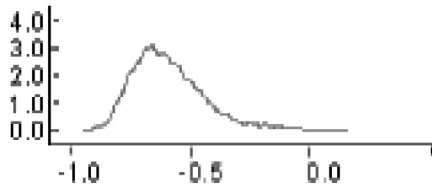
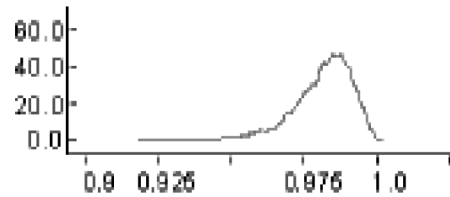
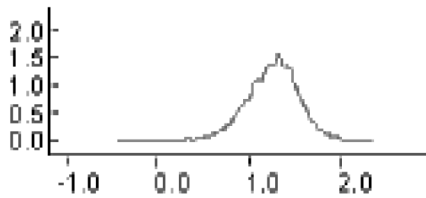
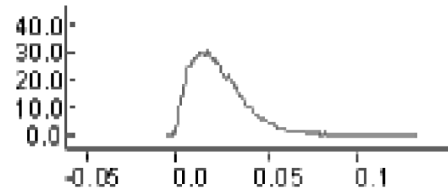
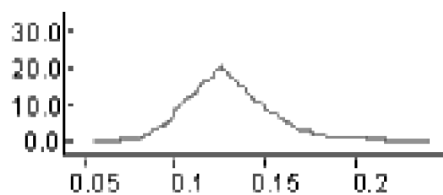
Figure 2. Density of parameter ρ Figure 3. Density of parameter ϕ Figure 4. Density of parameter μ Figure 5. Density of parameter α Figure 6. Density of parameter τ

Figure 7 presents the 95% confidence interval obtained for the volatility of the EUR/PLN. Similarly to the case of the WIG20, there is a pick at the end of 2008 and at the beginning of 2009. However, in the case of the WIG20, where the second pick is smaller, in the case of the EUR/PLN both of them reach approximately the same value.

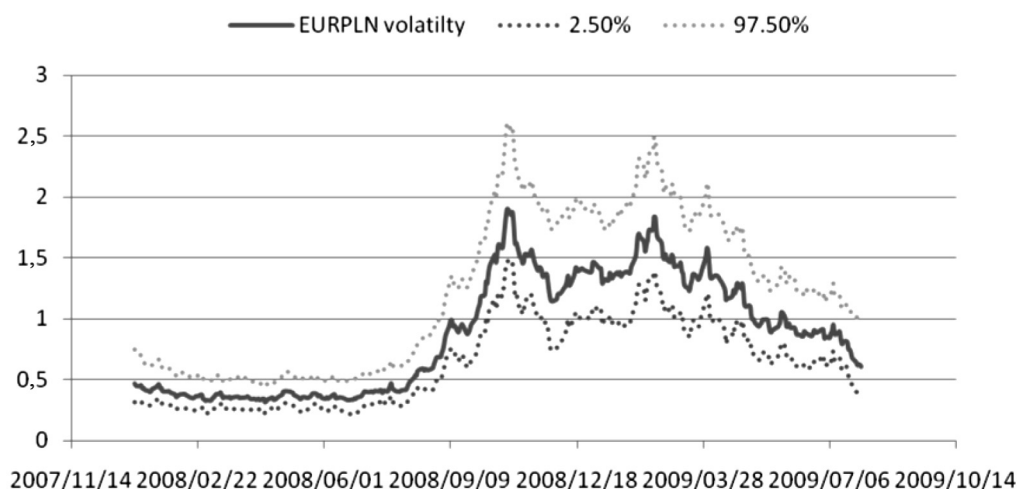


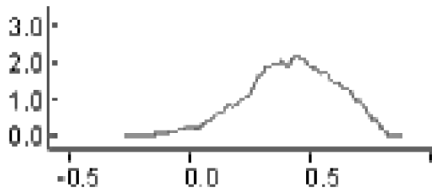
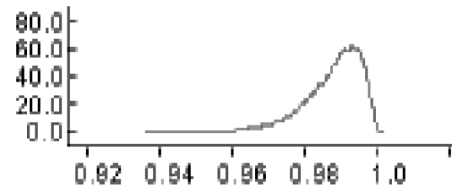
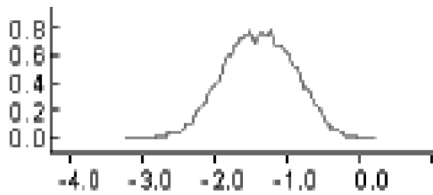
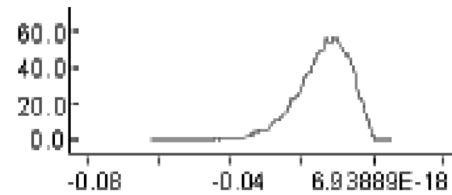
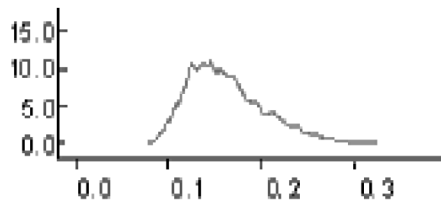
Figure 7. 95% confidence intervals obtained for volatility of the EUR/PLN exchange rat (leverage SV model)

In the case of the EUR/PLN exchange rate we do not observe the leverage effect, since the posterior mean of ρ takes positive value of 0.425. Also, if we take a look at its estimated density (Figure 8), we can observe that it is very unlikely that it will take a negative value. However, the value taken by the parameter ϕ indicates again the high persistence of volatility. Figures 8 to 12 present the densities obtained for each parameter, while their descriptive statistics are presented in Table 8.

Table 8

Descriptive statistics and 95% confidence interval for parameters of the EUR/PLN foreign exchange volatility

Parameter	Mean	Standard deviation	MC error	2.5%	Median	97.5%
τ	0.162	0.041	0.004	0.099	0.156	0.257
ρ	0.425	0.187	0.015	0.016	0.434	0.749
ϕ	0.988	0.008	0.001	0.968	0.990	0.999
α	-0.014	0.008	0.000	-0.033	-0.013	-0.002
α	-1.381	0.487	0.035	-2.331	-1.376	-0.467

Figure 8. Density of parameter ρ Figure 9. Density of parameter ϕ Figure 10. Density of parameter μ Figure 11. Density of parameter α Figure 12. Density of parameter τ

5. REALIZED VOLATILITY MEASURES AND CONDITIONAL VARIANCES FROM GARCH AND SV MODELS

5.1. REALIZED VOLATILITY IN DIFFERENT FREQUENCIES

In Figure 13 two kinds of realized volatility (with and without night return) and from different frequencies are presented, together with the estimated conditional variance from the GJR-GARCH and the SV models.

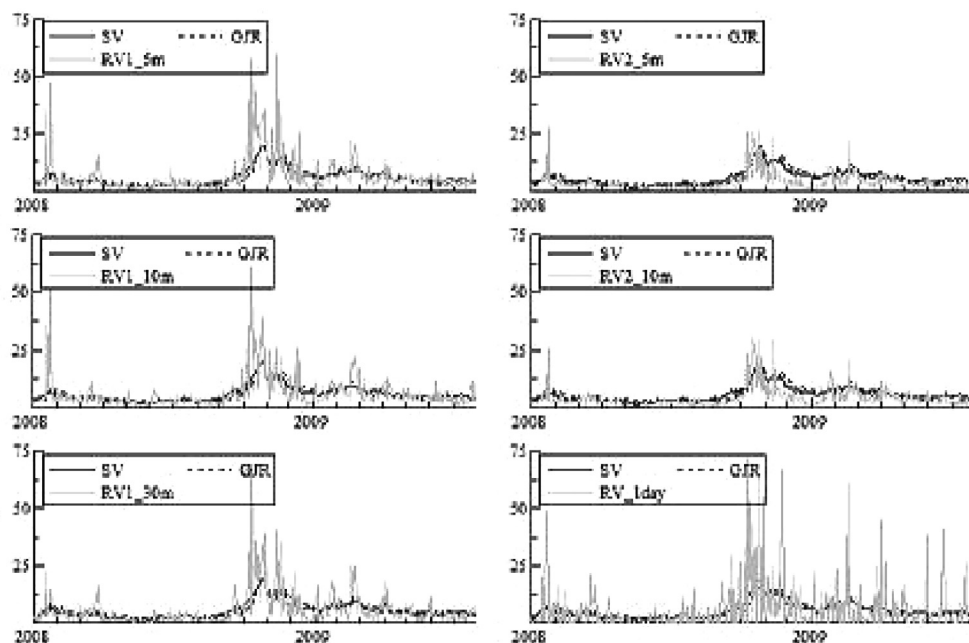


Figure 13. Realized volatility based on 5-, 10- and 30-minute returns with (RV1) and without (RV2) night return and square of daily returns (RV_1day) compared to estimates of conditional variance from the GJR-GARCH(1,1) and the SV models.

In Figure 13 realized volatility with night return (RV1) takes significantly higher values than realized volatility without night return (RV2). In other words, the realized volatility without night returns is much more smoothed. Such spectacular differences are not recognizable, however, if we compare realized volatility with night returns but from different frequencies – based on 5- and 10-minute returns (e.g. RV1_5m and RV1_10m). Eventually, if one would like to compare conditional variance from the GJR-GARCH and the SV models, the estimates of volatility are quite close to each other and the direction of changes (increase or decrease) in volatility is similar to that obtained within realized volatility.

5.2. COMPARISON OF REALIZED AND CONDITIONAL VARIANCE IN-SAMPLE (MINCER-ZARNOWITZ REGRESSION)

We compare the realized volatility measures among themselves. Because we do not know which approach, parametric or nonparametric, gives us the best estimates of unobserved volatility, we use the Mincer-Zarnowitz regression to our estimates. If the estimated parameters in this regression are as expected, it means α_0 close to 0, α_1 close to 1 and R^2 is not lower than 30%, this will indicate that the differences between the compared estimates are not huge and if so, both of them are good estimates of variance.

Table 9

Mincer-Zarnowitz in-sample regression for the WIG20 volatility

$$RV_t = \alpha_0 + \alpha_1 \hat{\sigma}_t^2 + u_t$$

	RV1			RV2			RV		
	α_0	α_1	R^2	α_0	α_1	R^2	α_0	α_1	R^2
	5 minutes			5 minutes			1 day		
GJR-GARCH(1,1)	-0,817	1,267	0,342	0,010	0,655	0,360	0,857	0,899	0,094
	(0,573)	(0,089)		(0,285)	(0,044)		(0,907)	(0,140)	
SV	-2,246	1,503	0,425	-0,602	0,771	0,440	-1,351	1,282	0,170
	(0,567)	(0,088)		(0,282)	(0,044)		(0,920)	(0,143)	
	10 minutes			10 minutes			30 minutes		
GJR-GARCH(1,1)	-0,969	1,301	0,323	-0,174	0,693	0,356	-0,953	0,527	0,350
	(0,613)	(0,095)		(0,304)	(0,047)		(1,188)	(0,0816)	
SV	-0,425	1,177	0,591	-0,218	0,838	0,723	-2,418	1,433	0,449
	(0,179)	(0,033)		(0,095)	(0,017)		(0,514)	(0,078)	

RV1 stands for realized volatility calculated with night returns and RV2 is calculated without night returns. 5, 10, and 30 minutes denote the frequencies of returns. RV 1 day stands for square of daily returns.

Parameters of the Mincer-Zarnowitz regression presented in Table 9 imply that the realized volatility is a better measure of unobserved volatility than the commonly used squares of returns (R^2 values are the lowest). When we compare realized volatility from different frequencies, in both cases: RV1 and RV2, estimates of conditional variance from the GJR-GARCH models are closer to realized volatility based on 5-minute returns, while estimates of conditional variance from the SV models are significantly closer to realized volatility based on 10-minute returns (and not 30-minute returns). When we compare two types of realized volatility, with (RV1) and without night return (RV2), the realized volatility without night return seems to perform better, especially for RV measure based on 10-minute returns.

In Table 10 the results of the Mincer-Zarnowitz in-sample regression for the EUR/PLN exchange rate are presented.

Table 10

Mincer-Zarnowitz regression in-sample for the EUR/PLN exchange rate volatility

	RV1			RV2			RV		
	α_0	α_1	R^2	α_0	α_1	R^2	α_0	α_1	R^2
	10 minutes			10 minutes			1 day		
IGARCH	0,120	1,083	0,453	0,061	0,948	0,508	0,163	0,858	0,148
	(0,084)	(0,060)		(0,065)	(0,046)		(0,145)	(0,103)	
SV	-0,102	1,347	0,618	-0,105	1,151	0,661	-0,161	1,222	0,2657
	(0,071)	(0,053)		(0,056)	(0,042)		(0,1369)	(0,102)	
	20 minutes			20 minutes			30 minutes		
IGARCH	0,104	1,099	0,451	0,085	1,056	0,454	0,105	1,125	0,448
	(0,086)	(0,061)		(0,082)	(0,059)		(0,088)	(0,062)	
SV	-0,113	1,357	0,606	-0,119	1,303	0,606	-0,125	1,399	0,612
	(0,074)	(0,055)		(0,071)	(0,053)		(0,075)	(0,056)	

The realized volatility of any frequency in intraday data is a better estimate of volatility than squares of daily returns (R^2 value is the lowest for the latter one). Similar to the WIG20 case, estimates of conditional variance from stochastic volatility models are closer to realized volatility than that from IGARCH model. When comparing realized volatility with and without night return, the latter is slightly closer to conditional variance estimates from parametric models. Moreover, in the case of the model for foreign exchange EUR/PLN the parameters estimated in the Mincer-Zarnowitz regression are as expected: α_0 is not statistically significantly different from zero and α_1 is very close to one.

6. THE ACCURACY OF GARCH AND SV FORECASTS – COMPARISON TO REALIZED VOLATILITY MEASURES

As a next step we computed 70 one-step ahead forecasts of the conditional variance within the GJR-GARCH (WIG20), the IGARCH (EUR/PLN) and the SV models (WIG20 and EUR/PLN). Next we compare them to different measures of realized volatility. In this section we answer the question: which frequency or approach of calculating realized volatility gives the non-parametric estimates of variance which are the closest to the variance forecast from the GARCH and the SV models. We take into account the following forecast error measures:

$$MAE = \frac{1}{T} \sum_{t=1}^T |\sigma_t^2 - RV_t|,$$

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (\sigma_t^2 - RV_t)^2},$$

$$MAPE = \frac{1}{T} \sum_{t=1}^T \left| \frac{\sigma_t^2 - RV_t}{\sigma_t^2} \right|,$$

$$AMAPE = \frac{1}{T} \sum_{t=1}^T \left| \frac{\sigma_t^2 - RV_t}{\sigma_t^2 + RV_t} \right|.$$

Forecast error measures from Mincer-Zarnowitz out-of-sample regression for WIG20 volatility are presented in Table 11.

Table 11

Forecasts error measures for differently calculated realized volatility of the WIG20 index

	MAE		RMSE		MAPE		AMAPE	
	GJR-GARCH (1,1)	SV	GJR-GARCH (1,1)	SV	GJR-GARCH (1,1)	SV	GJR-GARCH (1,1)	SV
RV1_5m	1.542	1.321	2.562	2.358	0.450	0.376	0.180	0.152
RV2_5m	1.355	1.152	1.602	1.326	0.376	0.336	0.239	0.215
RV1_10m	1.575	1.352	2.308	2.089	0.459	0.386	0.194	0.166
RV2_10m	1.531	1.347	1.793	1.521	0.429	0.393	0.277	0.257
RV1_30m	1.813	1.602	2.543	2.312	0.526	0.460	0.256	0.231
RV_1day	3.879	3.795	5.509	5.442	1.118	1.095	0.535	0.530

MAE stands for Mean Absolute Error; RMSE – Root Mean Squared Error; MAPE – Mean Absolute Percentage Error; AMAPE – Adjusted Mean Absolute Percentage Error.

We compare conditional variance forecasts to realized volatility calculated in different frequency (5, 10 and 30 minutes) and in a different way (with or without night returns). Generally, we conclude that the realized volatility from the highest examined frequency performs better than from the lower. It is in agreement with the results of Doman [10], where realized volatility from the highest frequency performs the best when forecast error measures are taken into account. However, in our study when comparing 5-minute realized volatility calculated with and without night return we do not have clear answer which of these two measures of realized volatility is better.

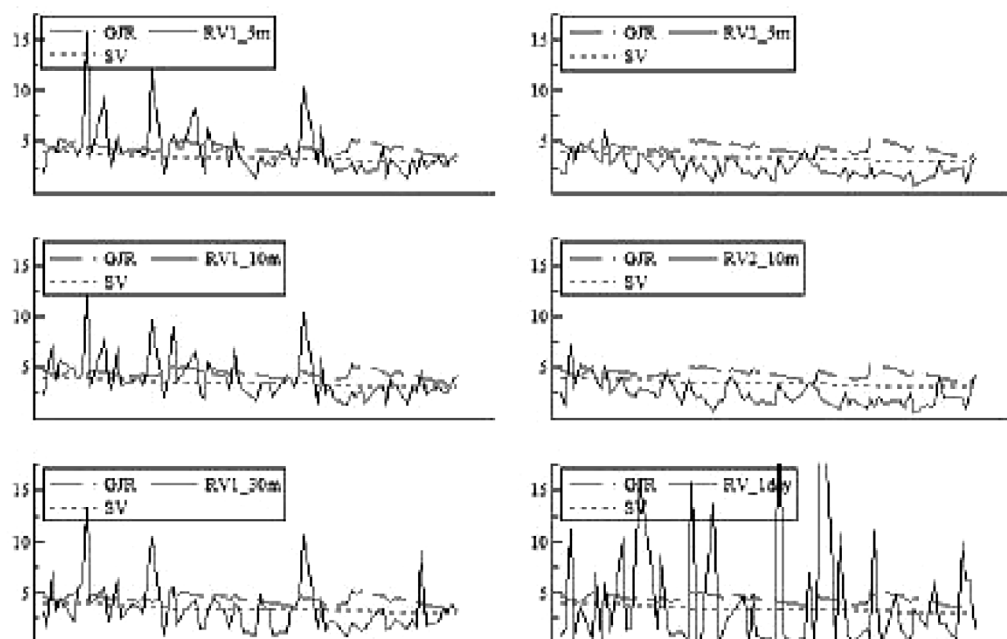


Figure 14. Conditional variance forecasts from GARCH and SV models with realized volatility of the WIG20 index

In Figure 14 we show conditional variance forecasts from the GJR-GARCH and SV models with realized volatility from different frequencies. The realized volatility with night return is generally higher than variance forecast, while realized volatility without night return is generally lower. When we look at the squared returns (realized volatility based on 1-day return), this measure is definitely the least adequate.

Table 12

Forecasts error measures for differently calculated realized volatility of the EUR/PLN exchange rate

	MAE		RMSE		MAPE		AMAPE	
	IGARCH(1,1)	SV	IGARCH(1,1)	SV	IGARCH(1,1)	SV	IGARCH(1,1)	SV
RV1_5m	0.253	0.281	0.325	0.381	0.435	0.810	0.206	0.254
RV2_5m	0.227	0.212	0.289	0.298	0.377	0.611	0.204	0.213
RV1_10m	0.284	0.299	0.379	0.437	0.491	0.864	0.229	0.259
RV2_10m	0.280	0.275	0.371	0.413	0.477	0.793	0.233	0.245
RV1_30m	0.739	0.922	1.136	1.292	1.305	2.630	0.335	0.488
RV_1day	0.464	0.416	0.625	0.634	0.800	1.191	0.479	0.491

We calculate forecast error measures having 70 one-step ahead forecasts from IGARCH model for the EUR/PLN exchange rate. The error measures presented in Table 12 show that the lowest values of the errors are – again – obtained for the highest frequency, which is 10-minutes in this case. The conditional variance from the IGARCH and SV models together with differently calculated realized volatility are presented in Figure 15.

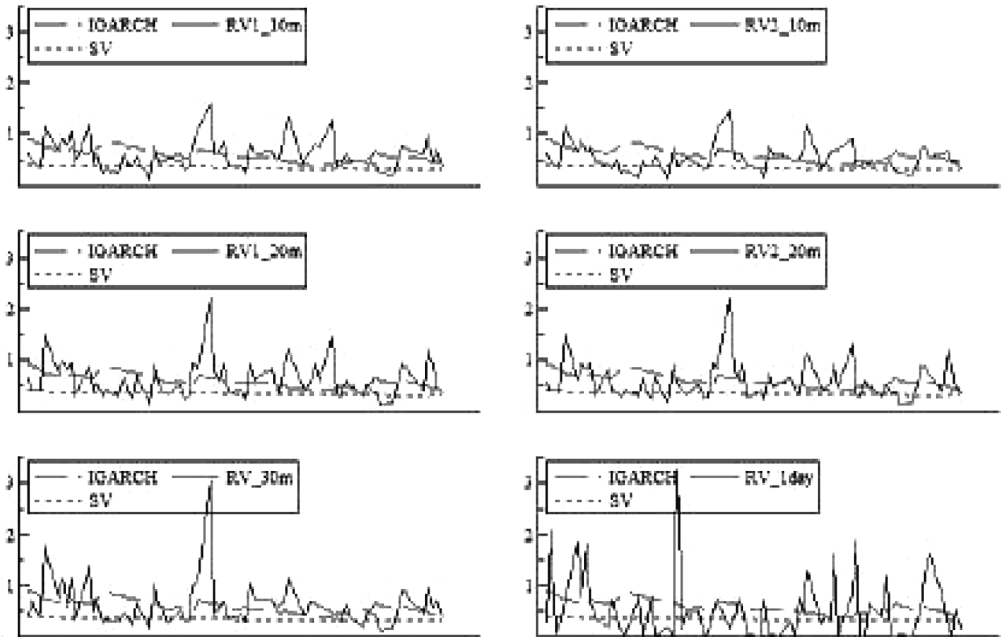


Figure 15. Conditional variance forecasts and realized volatility of the EUR/PLN exchange rate

7. CONCLUSIONS

The aim of the paper was to compare three types of volatility estimates: GARCH, SV and RV for the given instruments on the Polish financial markets: the stock exchange and the foreign currency ones. Based upon our samples: the WIG20 index and the EUR/PLN foreign exchange rate, both covering the crisis period, we computed the mentioned measures of volatility and compared them using the simplest and most commonly applied technique: the Minzer-Zarnowitz regression. Contrary to most commonly applied procedure, we used not only the estimates of RV to compare SV and GARCH models, but we also compared the RV measures among themselves. Since volatility is not observable, we cannot decide estimates of which models are the best measures of volatility of a financial instrument. Thus, we assume that two measures of volatility are equally good if the R^2 coefficient of the regression is quite high, the

parameter α_0 is close to 0 and α_1 to 1. This assumption allowed us to compare the RV measures obtained from the data of different frequencies.

In the case of the WIG20 index and the EUR/PLN exchange rate it emerged that the square of daily returns is the worst estimate of volatility from all applied RV models. In both cases RV computed based upon 10-minute returns without the night return performed the best (according to R^2 measure). Moreover, in the case of the WIG20 index the conditional variance from GJR-GARCH model was closer to the RV computed based upon 5-minute returns, while the Leverage-SV model – to the one computed based upon the 10-minute returns. However, in case of the EUR/PLN exchange rate both the SV and the IGARCH models were close to the RV computed based upon 10-minute returns.

Eventually, we compared the RV estimates with conditional variance from the GARCH and the SV models estimated for the WIG20 index and the EUR/PLN foreign exchange. In general, it seems that RV computed from data of higher frequency performed better than others. The same conclusion can be drawn in the case of the EUR/PLN exchange rate, although here the preferred model was the one without the night return (in the case of the WIG20 index the results were ambiguous).

Poznań University of Economics

REFERENCES

- [1] Akaike H., [1973], *Information theory and an extension of the maximum likelihood principle*, [in:] B. Petrov, F. Csaki, (eds.), *Proceedings of 2nd International Symposium on Information Theory*, Academia Kiado, Budapest, 267-281.
- [2] Andersen T., Bollerslev T., Diebold F., [2003], *Parametric and nonparametric volatility measurement*, [in:] L.P. Hansen, Y. Ait-Sahalia (eds.), *Handbook of Financial Econometrics*, Amsterdam, North-Holland.
- [3] Black F., [1976], *Studies of Stock Price Volatility Changes*, *Proceedings of the 1976 Meeting of the American Statistical Association*, 177-181.
- [4] Bollerslev T., [1986], *Generalized autoregressive conditional heteroscedasticity*, „*Journal of Econometrics*” 31, 307-327.
- [5] Bollerslev T., Chou R.Y., Kroner K.F., [1992], *ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence*, „*Journal of Econometrics*” 52, 5-59.
- [6] Broto C., Ruiz E., [2004], *Estimation Methods for Stochastic Volatility Models: Survey*, „*Journal of Economic Surveys*” 18, 613-649.
- [7] Clark P.K., [1973], *A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices*, „*Econometrica*” 41, 135-155.
- [8] Doman M., Doman R., [2004], *Stochastic Volatility Models of Exchange Rates in Central European Countries: effects of Exchange Rate Regimes*, „*Przegląd Statystyczny*” 51 (4), 37-55.
- [9] Doman M., Doman R., [2009], *Modelowanie zmienności i ryzyka (Volatility and Risk Modeling)*, Oficyna Wydawnicza Wolters Kluwer Polska, Kraków.
- [10] Doman R., [2003], *Prognozowanie zmienności zrealizowanej indeksu WIG20 za pomocą modelu GARCH (Forecasting the Realized Volatility of the WIG20 index Using the GARCH model)*, „*Przegląd Statystyczny*” 50 (4), 147-163.
- [11] Engle R.F., [1982], *Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation*, „*Econometrica*” 50, 987-1007.
- [12] Glosten L.R., Jagannathan R., Runkle D., [1993], *On the Relation Between Expected Value and the Volatility of the Nominal Excess Return on Stocks*, „*Journal of Finance*” 48, 1779-1801.

- [13] Harvey A.C., Shephard N., [1996], *The Estimation of an Assymetric Stochastic Volatility Models for Asset Reurns*, „Journal of Business and Economic Statistics” 14, 429-434.
- [14] Jacquier E., Polson N., Rossi P., [2004], *Bayesian analysis of stochastic volatility models with fat-tails and correlated errors*, „Journal of Econometrics”, 122, 185-212.
- [15] Kim S., Shephard N., Chib S., [1998], *Stochastic volatility: likelihood inference and comparison with ARCH models*, „Review of Economic Studies” 65, 361-393.
- [16] Laurent S., Peters J.P., [2002], G@RCH 2.2: An Ox Package for Estimating and Forecasting Various ARCH Models, „Journal of Economic Surveys”, 16, 447-485.
- [17] Mandelbrot B., [1963], *The Variation of Certain Speculative Prices*, „Journal of Business”, 36, 394-419.
- [18] Mele A., Formari F., [2000], *Stochastic Volatility in Financial Markets: Crossing the Bridge to Continuous Time*, Kluwer Academic Publishers.
- [19] Mincer J., Zarnowitz V., [1969], *The Evaluation of Economic Forecasts and Expectations*, National Bureau of Economic Research.
- [20] Ntzoufras I., [2009], *Bayesian Modeling using WinBUGS*, John Wiley and Sons.
- [21] Pajor A., [2003], *Procesy zmienności stochastycznej SV w bayesowskiej analizie finansowych szeregów czasowych (Stochastic Volatility Processes In Bayesian Analysis of Financial Time Series)*, doctoral dissertation (In Polish), published by Cracow University of Economics, Kraków.
- [22] Osiewalski J., Pajor A., Pipień M., [2006], *Bayesian Analysis of Main Bivariate GARCH and SV Models for PLN/USD and PLN/DEM (1966-2001)*, [in:] Zieliński Z. (ed.), *Dynamic Econometric Models 7*, ISSN: 12343862.
- [23] Shepard N., [1996], *Statistical Aspects of ARCH and Stochastic Volatility*, [in:] *Time Series Models in Econometrics, Finance and Other Fields*, ed. Cox D.R., Hinkley D.V., Barndorff-Nielsen O.E., Champan and Hall, London.
- [24] Speighalter D., Best N., Carlin B., van der Linde A., [2002], *Bayesian measures of model complexity and fit (with discussion)*, *Journal of the Royal Society B* 64, 583-639.
- [25] Taylor S., [1986], *Modeling Financial Time Series*, John Wiley and Sons.
- [26] Yu J., [2005], *On Leverage in a Stochastic Volatility Model*, „Journal of Econometrics”, 127, 165-178.
- [27] Yu J., Meyer R., [2000], *BUGS for a Bayesian analysis of Stochastic Volatility models*, „Econometric Journal” 3, 198-215.
- [28] <http://www.mysmu.edu/faculty/yujun/research.html>

Praca wpłynęła do redakcji w marcu 2010 r.

ZMIENNOŚĆ ZREALIZOWANA WOBEC MODELI GARCH I MODELI ZMIENNOŚCI STOCHASTYCZNEJ NA POLSKIM RYNKU KAPITAŁOWYM

Streszczenie

Celem artykułu jest porównanie oszacowań zmienności uzyskanych z modeli parametrycznych: GARCH i SV z oszacowaniem uzyskanym na podstawie zmienności zrealizowanej szacowanej w oparciu o dane różnej częstotliwości. W badaniu wzięto pod uwagę zwroty z wybranych instrumentów polskiego rynku finansowego: indeks WIG 20 oraz kurs walutowy EUR/PLN. Ujęta w badaniu próba objęła okres kryzysu finansowego, co stanowi istotne uzupełnienie wyników prezentowanych do tej pory w literaturze.

Słowa kluczowe: zmienność zrealizowana, SV, GARCH, prognozowanie zmienności

REALIZED VOLATILITY VERSUS GARCH AND STOCHASTIC VOLATILITY MODELS.
THE EVIDENCE FROM THE WIG20 INDEX AND THE EUR/PLN FOREIGN EXCHANGE MARKET

Summary

The aim of the article is to compare the estimates of the volatility obtained from the parametric models: the GARCH and the SV with the estimates based upon the Realized Volatility approach, whereas the estimates from the RV are obtained from the data of different frequencies. The data sample consists of the WIG20 index and the EUR/PLN exchange rate and covers the hectic crisis period. Hence, the presented results can be viewed as an extension of the results of the studies presented up to date.

Key words: realized volatility, SV, GARCH, volatility forecasting

ANNA CZAPKIEWICZ, IWONA SKALNA

THE CAPM AND FAMA-FRENCH MODELS IN POLAND

1. INTRODUCTION

The cross-sectional differences in asset expected returns have attracted considerable attention in finance literature. The Single-Factor Capital Asset Pricing Model (CAPM) and the Fama-French Three-Factor (FF) models occupy an essential place among models of expected returns.

The CAPM model of William Sharpe [28] and John Lintner [21], although simple and logical, is based on certain restrictive assumptions about the functioning of the market. By the end of 70's and early 80's, the empirical evidence against the CAPM became very strong ([2], [3]). It was noticed, that return differences of some assets grouped according to the firms' financial characteristics were not captured by the betas. Some attempts to explain these anomalies were made, among others, by Banz [2] and Reinganum [25] who investigated the size effect and found a return premium on some small stocks. The book-to-market effect was researched by Rosenberg et al. [26] and confirmed later by Blume and Stambaugh [6]. They found the premium in the case of stocks with high ratios. Fama and French [10] had investigated the explanatory power of returns on some factors associated with the companies' characteristics, such as size, book value to market value ratio, leverage stock's price earnings ratio. As a result of their efforts, they proposed a three factors model to explain the stock returns: the excess of return in relation to the market (market factor), the difference between the returns of portfolio with the large and small capitalization (size factor, SMB) and the difference between the returns on portfolios of high and low book value to market value (HML). The inclusion of those factors improved the model fitting to the US empirical data. The investigations conducted for other markets (Australian, Canadian, German, French, Japan, and UK) showed that the size and BE/ME effects might have an international character. The asset pricing model on Polish stocks has been examined and tested, among others, by Bołt and Miłobędzki [7], Fiszeder [12], Kowerski [18], Urbański [30], Zarzecki [31].

In this study, we investigate the ability of the Fama-French three-factor model to explain cross-sectional differences in asset returns for the period December 2002 to January 2010. The research is conducted on the Warsaw Stock Exchange (WSE). The financial data originate from the CEDULA¹ which is a daily official bulletin providing information from WSE. The systematic risk and the risk premium estimations

¹ www.gpw.com.pl

are estimated using the Generalized Method of Moments (GMM) which takes into account autocorrelation and heteroscedasticity in time series. Procedures for testing whether the intercepts are jointly null is carried out by test proposed by MacKinlay and Richardson [22]. All calculations presented in this study have been performed using application developed and dedicated for the purposes of this paper.

The study has the following structure. In Section 2, the econometric model and the methodology for estimating model parameters are presented. The empirical study is discussed in Section 3. The paper ends with concluding remarks.

2. THEORY AND METHODOLOGY

According to a K factor asset pricing model, the expected return on a portfolio in excess of the risk-free rate is explained by its sensitivity of its return to K common factors. The expected excess returns satisfy the linear relations which can be written in a matrix form as:

$$E(\mathbf{R}_t) = \gamma_0 + \gamma_1\beta_1 + \dots + \gamma_K\beta_K, \quad (1)$$

where $\mathbf{R}_t$ is an $N \times 1$ vector of portfolios' excess returns at time t

$$\mathbf{R}_t = \alpha + \beta_1 f_{1t} + \dots + \beta_K f_{Kt} + \varepsilon_t, \quad t = 1, \dots, T, \quad (2)$$

$\gamma_1, \dots, \gamma_K$ are risk premiums, $\alpha, \beta_1, \dots, \beta_K$ are $N \times 1$ vectors of factor sensitivities or loadings, $f_{1t}, \dots, f_{Kt}$ are common factors at time t , and ε_t is an $N \times 1$ vector of error terms at time t .

2.1. THE SYSTEMATIC RISK ESTIMATION AND THE INTERCEPT TEST

The estimation of the systematic risk parameters is performed using the Generalized Method of Moments [15] approach which is robust to both conditional heteroscedasticity and serial correlation in the return residuals as well as in the factors.

The GMM estimator of the vector of parameters $\varphi = (\alpha^T, \beta_1^T, \dots, \beta_T^T)^T$ must fulfill the following conditions:

$$E[g_t(\varphi)] = E\left[\varepsilon_t \otimes \begin{pmatrix} 1 \\ F_t \end{pmatrix}\right] = E\left[\begin{pmatrix} R_t - \alpha - \beta F_t \\ (R_t - \alpha - \beta F_t) \otimes F_t \end{pmatrix}\right] = \mathbf{0} \quad (3)$$

where $F_t = (f_{1t}, \dots, f_{Kt})^T$ is the $K \times 1$ vector of factors at time t , and $\beta = [\beta_1, \dots, \beta_K]$ is the $N \times K$ matrix of unknown parameters. If the system is exactly identified, i.e. the number of equations is equal to the number of unknowns, the GMM procedure is equivalent to the Ordinary Least Square (OLS) regression.

Hansen [15] shows that the GMM estimator $\hat{\varphi}$ has an asymptotic normal distribution with the mean equal to φ and asymptotic variance matrix

$$\text{var}[\hat{\varphi}] = (\mathbf{D}^T \mathbf{S}_0^{-1} \mathbf{D})^{-1} \quad (4)$$

where $\mathbf{D}$ is the matrix of the derivatives of $g_t(\varphi)$ with respect to all parameters, and

$$\mathbf{S}_0 = \sum_{j=-\infty}^{\infty} E[g_t(\varphi) g_{t-j}^T(\varphi)].$$

To account for heteroscedasticity and autocorrelation, the consistent estimator $\mathbf{S}_T$ of $\mathbf{S}_0$ can be defined as

$$\mathbf{S}_T = \mathbf{\Omega}_0 + \sum_{j=1}^m w(j, m) [\mathbf{\Omega}_j + \mathbf{\Omega}_j^T],$$

where $w(j, m)$ is some weighting function, and

$$\mathbf{\Omega}_j = \frac{1}{T} \sum_{t=j+1}^T g_t(\hat{\varphi}) g_{t-j}^T(\hat{\varphi}), \quad j = 0, \dots, m.$$

(assuming that $E(f_i \varepsilon_i) = 0$ and $m = 0$, variances of $\hat{\varphi}$ are equal to variances obtained using the separate OLS regressions for each portfolio).

The multifactor model can generate efficient portfolios if all intercepts are jointly equal to zero. An ordinary (GRS) test for that implication was suggested by Gibbons, Ross and Shanken [13]. It works correctly, when the disturbances are temporally independent and jointly normal, with the zero mean. This test, however, does not adjust for autocorrelated series, which is a feature often found in financial series. The test based on GMM approach, proposed by MacKinlay and Richardson [22], can be employed to test null hypothesis that the intercept vector is zero. In this case, the test statistic can take the following form:

$$\chi_N^2 = \hat{\alpha}' \text{var}[\hat{\alpha}]^{-1} \hat{\alpha}, \quad (5)$$

where $\text{var}[\hat{\alpha}]$ is the upper left corner of matrix (4). In case the null hypothesis is true, this statistic is distributed as χ^2 with N degrees of freedom. To improve the finite sample behavior of this test, this statistic should be scaled by $\left(\frac{T-N-K}{T}\right)$ ([22]).

2.2. THE RISK PREMIUM ESTIMATION

The risk-return relationship can be estimated in two stages. At the first stage, beta estimates are obtained from the separate time-series regressions for each asset (2).

Then, these estimates are used in the second-pass cross-sectional regression (CSR) (1). Since the independent variable in the CSR is measured with an error, the second-pass estimate is a subject to an errors-in-variables problem. To omit this problem, all unknown parameters $\varphi = (\alpha^T, \beta_1^T, \dots, \beta_K^T, \gamma_0, \dots, \gamma_K)^T$ in (1) and (2) should be estimated simultaneously. The GMM approach can be used for this purpose since it is robust to both the conditional heteroscedasticity and serial correlation in the return residuals as well as in the factors. However, it is difficult to obtain the GMM estimators. This is due to a large number of parameters to estimate coupled with the nonlinearity of the model [27].

In this work we consider the two step GMM approach suggested by Harvey and Kirby [16]. The model is estimated sequentially. First, the parameter vector $\varphi = (\alpha^T, \beta_1^T, \dots, \beta_K^T, \gamma_0, \dots, \gamma_K)^T$ is partitioned into two sub-vectors $\varphi = (\varphi_1^T, \varphi_2^T)^T$, where $\varphi_1 = (\alpha^T, \beta_1^T, \dots, \beta_K^T)^T$ and $\varphi_2 = (\gamma_0, \dots, \gamma_K)^T$. Then, the sample moments can be expressed as:

$$g_t(\varphi) = [\varepsilon_t(\varphi_1) \otimes F_t^T, (R_t - [1_N, \beta]\varphi_2)^T]^T = [g_{1t}(\varphi_1), g_{2t}(\varphi)]^T$$

and the GMM conditions are:

$$E[g_{1t}(\varphi_1), g_{2t}(\varphi)]^T = \mathbf{0}. \quad (6)$$

The first part of the GMM condition is exactly identified. Hence, the estimator $\hat{\varphi}_1$ is equal to the time-series OLS estimator obtained from equation (2). The two step strategy of the GMM method uses the estimated parameters $\hat{\varphi}_1$ in the second part of condition (6). Hence, the problem boils down to find a minimum

$$\min_{\varphi_2} g_{2T}(\hat{\varphi}_1, \varphi_2)^T \mathbf{W}_{2T} g_{2T}(\hat{\varphi}_1, \varphi_2),$$

where $\mathbf{W}_{2T}$ is a weighting matrix. Following Ogaki [23], the optimal weighting matrix can be chosen as

$$\mathbf{W}_{2T} = [[-D_{21} D_{11}^{-1} I_N] S_T [[-D_{21} D_{11}^{-1} I_N]^T]^{-1} \quad (7)$$

where $\mathbf{D}_{ij}$, $i, j = 1, 2$, is the (i, j) block of the $N(K+2) \times (N+1)(K+1)$ matrix of the derivatives of $g_T(\varphi)$ with respect to all the parameters, and $\mathbf{S}_T$ is a consistent estimator of the covariance matrix of moment conditions (6). Since $\mathbf{D}_{21}$ in (7) is a zero matrix, thus the weighting $\mathbf{W}_{2T}$ matrix reduces to

$$\mathbf{W}_{2T} = [[0 \ I_N] S_T [[0 \ I_N]^T]^{-1}.$$

The above discussion leads to the conclusion, that the consistent GMM estimator $\hat{\varphi}_2$ satisfies the minimum of the function:

$$(\bar{R} - [\mathbf{I}_N, \hat{\beta}] \varphi_2)^T S_{2T}^{-1} (\bar{R} - [\mathbf{1}_N, \hat{\beta}] \varphi_2),$$

where $\bar{R}$ is a vector of the averages of the excess returns of the portfolios, and the covariance matrix S_{2T} is estimated following Andrews [1]:

$$S_{2T} = \Omega_0 + \sum_{j=1}^m w(j, m) [\Omega_j + \Omega_j^T],$$

with $w(j, m)$ being Parzen or Bartlett kernel and

$$\Omega_0 = \frac{1}{T} \sum_{t=j+1}^T \mathbf{h}_t \mathbf{h}_t^T, \quad \mathbf{h}_t = \hat{\beta}(F_t - \bar{F}) + \varepsilon_t,$$

where $\bar{F}$ is a vector of factor averages.

The obtained $\hat{\varphi}_2$ estimator is used as a starting point for the iteration to attain the efficiency bound to the GMM estimator. In this work, the second stage is iterated with the inverse of the long run covariance matrix until the convergence condition is fulfilled.

The advantage of the two-step GMM procedure is that all estimations are linear, which makes it easier to find out the solution.

3. EMPIRICAL ANALYSIS

3.1. DATA SET AND PORTFOLIO CONSTRUCTION

This study uses the data set containing stock prices traded in the Warsaw Stock Exchange (the number of stocks varies from 120 to 400 during the sample period). The sample used spans from December 2002 to January 2010. All prices are closing prices adjusted for splits and dividends. The analysis is restricted to the companies with available returns. The risk free rate (R_f) is calculated from the 52-weeks treasury bills.

The sample is divided into 12 portfolios on the basis of the selected characteristics: size and BE/ME ratio. The portfolios are formed as follows. At the beginning of each month stocks are allocated to the four groups based on their sorted sizes. Next, the stocks in each group are allocated in an independent sort to three book-to-market equity (BE/ME) groups. The portfolios are updated at the beginning of each month. The averages and standard deviations of the excess return for the 12 size-BE/ME relative portfolios are reported in Table 1. The statistics show that the portfolios with a small capitalization and high BE/ME earn the highest average excess return but consist of the firms of the highest risk. This portfolio has the biggest average monthly excess return equal to 2.2% with the standard deviation of 11%, whereas the portfolio of big firms with low BE/ME has the average excess return of -1% with the smallest standard deviation equal to 7.4%.

Table 1

Descriptive statistics for 12 size-BE/ME portfolios for the sample period December 2002 to January 2010

Size	Summary statistics					
	Low	2	High	Low	2	High
	R_i					
	Average			Standard deviation		
Small	-0.013	0.012	0.022	0.109	0.111	0.110
2	-0.008	-0.001	-0.002	0.080	0.082	0.101
3	-0.006	-0.006	0.000	0.080	0.074	0.091
Big	-0.010	-0.015	0.000	0.074	0.092	0.081
	ln (ME)					
	Average			Standard deviation		
Small	3.426	3.350	3.237	0.518	0.501	0.627
2	4.549	4.521	4.528	0.425	0.455	0.485
3	5.776	5.677	5.640	0.409	0.415	0.421
Big	7.519	7.874	7.731	0.292	0.415	0.441
	ln (BE/ME)					
	Average			Standard deviations		
Small	-1.158	-0.176	0.446	0.733	0.617	0.577
2	-1.201	-0.443	0.190	0.492	0.471	0.458
3	-1.422	-0.701	-0.009	0.386	0.436	0.437
Big	-1.573	-0.851	-0.200	0.349	0.389	0.320

The monthly returns of the two selected portfolios are presented in Fig. 1. Portfolio 3 consists of small firms with the high BE/ME ratio, whereas Portfolio 10 consists of big firms with the low BE/ME. One can notice that monthly returns of portfolio 3 are higher than monthly returns of portfolio 10 in contrary to the deviations of the excess returns.

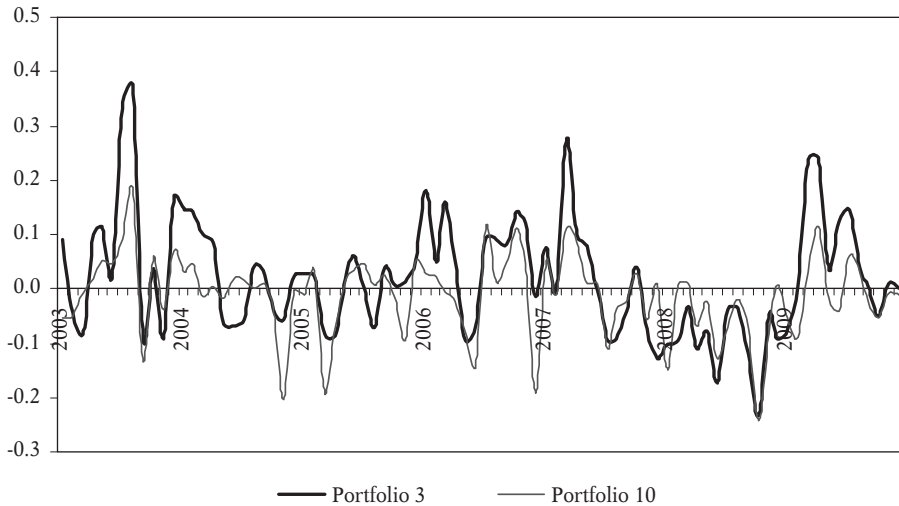


Figure 1. Monthly returns of the selected portfolios

The explanatory returns SMB, and HML are formed similarly to the 12 size-BE/ME relative portfolios. At the beginning of each month stocks are allocated to the two groups (S small or B big) based on the fact whether their market equity is below or above the median. Next, the stocks in each group are allocated in an independent sort to three book-to-market equity (BE/ME) groups (L low, M medium, or H high) based on the breakpoints for the bottom 30 percent, middle 40 percent, and top 30 percent of the values of BE/ME for WSE stocks. The SMB is the difference, each month, between the average of the returns on the three small-stock portfolios (S/L, S/M, and S/H) and the average of the returns on the three big-stock portfolios (B/L, B/M, and B/H). The HML is the difference between the average of the returns on the two high-BE/ME portfolios (S/H and B/H) and the average of the returns on the two low-BE/ME portfolios (S/L and B/L). The third factor (R_M) is the return of the market portfolio (WIG) in the excess of the risk-free rate. The time series of the factors are depicted in Figure 2. The sensitive similarity of the market and the HML factors can be observed.

Table 2

Summary statistics of the factors

	R_M	SMB	HML
Average	0.005	0.007	0.016
Standard deviation	0.072	0.037	0.046

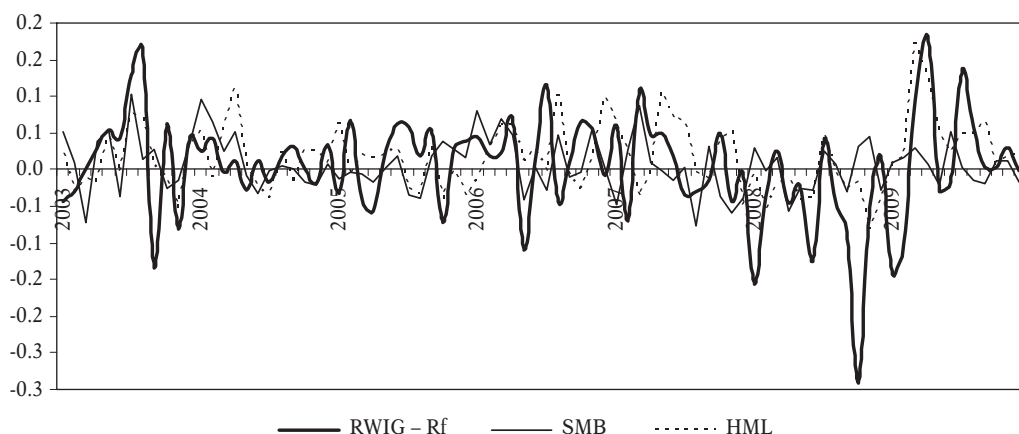


Figure 2. Time series of the factors

Table 2 reports the basic statistics of the factors. The mean of R_M is statistically insignificantly positive ($|t| = 0.69$), whereas the means of the SMB and HML are significantly positive $|t_{SMB}| = 1.65$ and $|t_{HML}| = 3.25$. This means that small size portfolios with high BE/ME gain bigger profits. These calculations confirm the basic model assumption, that the rate of the return depends on the risk connected with the investments in stocks of small firms and in the stocks of firms with high ratio of book value to market value which are underpriced by market.

Table 3

The correlation between factors; the corresponding p -values are given in parentheses

	R_M	SMB	HML
R_M	1		
SMB	0.136 (0.190)	1	
HML	0.331 (0.000)	0.192 (0.056)	1

It is desirable that the factors are not correlated. Table 3 shows that the SMB and HML factors are uncorrelated, however, a small (0.331) but significant ($p = 0.000$) correlation between the R_M and the HML factors can be noticed.

3.2. SYSTEMATIC RISK COMPONENTS

This section reports the estimators of the factor loadings calculated by regression (2) and the tests of the significance of the constant. These estimators are obtained

using the method described in Section 2.1. The variances of unknown parameters are obtained from (4) and calculated with the different bandwidths. The point estimators of unknown parameters in (2) are independent on the bandwidths in contrary to the variances. However, only a small change of the significance can be noticed. Therefore, sole zero bandwidth results are reported in Table 4.

Table 4

Time Series Regression of Portfolios

Regression: $R_{it} = \alpha_i + \beta_{R_M,i}R_{Mt} + \beta_{SMB,i}SMB_t + \beta_{HML,i}HML_t + \varepsilon_{it}$						
Size	Low	2	High	Low	2	High
	$\hat{\alpha}$			$t(\hat{\alpha})$		
Small	-0.019	-0.005	0.000	-3.068	-0.731	0.063
2	-0.011	-0.013	-0.021	-2.623	-2.375	-3.375
3	-0.007	-0.014	-0.015	-1.204	-2.822	-2.801
Big	-0.011	-0.020	-0.009	-2.379	-3.303	-2.692
	$\hat{\beta}_{R_M}$			$t(\hat{\beta}_{R_M})$		
Small	1.010	0.970	0.964	8.344	10.435	11.801
2	0.858	0.787	0.930	16.278	11.323	16.952
3	0.911	0.767	0.948	13.612	13.051	13.182
Big	0.858	1.069	0.968	13.561	16.795	14.150
	$\hat{\beta}_{SMB}$			$t(\hat{\beta}_{SMB})$		
Small	1.570	1.460	1.323	8.810	8.644	10.797
2	0.885	0.744	0.909	7.045	5.878	4.986
3	0.136	0.365	0.309	0.745	3.061	2.453
Big	0.115	0.170	-0.079	1.295	0.847	-0.901
	$\hat{\beta}_{HML}$			$t(\hat{\beta}_{HML})$		
Small	-0.578	0.114	0.435	-3.329	0.645	3.841
2	-0.426	0.140	0.444	-3.175	1.005	3.363
3	-0.310	0.094	0.482	-2.273	0.767	3.073
Big	-0.311	-0.127	0.288	-2.510	-0.894	3.178

The results of the estimation of the systematic risk vector are similar to Fama-French study of the American market in the years 1963-2001 (see [24]). We can notice that the HML loadings are significantly negative for the portfolios with low BE/ME values and increase monotonically in the significantly positive values for portfolios with high BE/ME values. The increase is independent on the size of the portfolios. These changes occur for all portfolios sorted by the firm size. The SMB loadings are significant for the small size portfolios and insignificant for big portfolios. We can also notice that this values monotonically decrease from portfolios with low BE/ME to portfolios with high BE/ME values. This regularity is noticed only for small portfolios. We can notice that the systematic risk connected with the market factor is independent on the method of constructing the portfolios. Hence, this factor shows a weak ability to explain the rate of returns.

Next, the tests whether the intercepts of equation (2) are jointly equal to zero, are carried out. The χ^2 statistic (5) is calculated for this purpose. Table 5 presents results for the one factor model (CAPM) and FF three-factor model for the same sample period. One can observe that the Fama-French three-factor model with additional loadings, the SMB and HML factors, is consistently better than the CAPM.

Table 5

Intercept tests: GRS and χ^2 and the corresponding p -values

CAPM		FF	
GRS	χ^2	GRS	χ^2
3.38 (0.001)	46.67 (0.000)	2.27 (0.017)	29.00 (0.004)

(Assuming that $E(f_i \varepsilon_i) = 0$, the GRS test and χ^2 test are asymptotically comparable.) The results indicate that FF model is better than the CAPM. However, one may argue that the three-factor model does not describe satisfactorily the rate of portfolios. Using the GRS test, we have obtained the GRS-statistic equal to 2.27 (for comparison, 2.91 were obtained for Fama-French 25 portfolios of US market [24]). This means that the Fama-French factors do not account for all factors describing the excess returns. The problem of the intercept test needs additional discussion which is out of scope of this study and will be the subject of the future work.

3.4. RISK PREMIUM COMPONENTS

This section presents the estimates of equation (1) obtained by the GMM method described in Section 2. The choice of the bandwidth is more critical than the choice of the weighting matrix [14]. Therefore, we estimate the model with three different bandwidths: zero lag, one lag and four lags. The latter is suggested by Hall [14] ($T^{1/3} \approx 4$)

because this minimizes the mean square error of the covariance estimator. We also set the bandwidth equal to one because the analysis of the series has shown that for most portfolios we can notice only one lag autocorrelation.

Table 6 presents the risk premium estimation and t -statistics for the twelve size-BE/ME portfolios in the FF three-factor model. The result of the CAPM approach are presented for the comparison purposes. It has been done only for GMM 0 lag.

Table 6

The risk premium estimation and t -statistics for portfolios based on ME and BE/ME criteria in the CAPM and FF three-factor model

Method	kernel	$\hat{\gamma}_0$	$t(\hat{\gamma}_0)$	$\hat{\gamma}_{R_M}$	$t(\hat{\gamma}_{R_M})$	$\hat{\gamma}_{SMB}$	$t(\hat{\gamma}_{SMB})$	$\hat{\gamma}_{HML}$	$t(\hat{\gamma}_{HML})$
GMM 0 lag		-0.042	-3.155	0.039	2.349				
		-0.017	-1.130	0.011	0.594	0.007	1.853	0.018	3.421
GMM 1 lags	Parzen	-0.019	-1.267	0.014	0.781	0.007	1.805	0.019	3.858
	Bartlett	-0.020	-1.364	0.019	1.035	0.007	1.810	0.021	4.228
GMM 4 lags	Parzen	-0.012	-0.839	0.014	0.783	0.009	2.369	0.022	4.441
	Bartlett	-0.007	-0.456	0.009	0.482	0.010	2.490	0.022	4.367

We have obtained significantly positive risk premium for SMB and HML factors. The risk premium connected with HML factor equals 2% and is higher than estimated risk premium connected with the firm size factor. More attention should be paid to the fact that the methods proved the insignificance of the constant. This seems to contradict with significantly different from zero the overall value of constants estimated in the first step. This problem requires further investigations. Moreover, some positive although insignificant value of the market risk factor can be noticed. This remarks implies that market beta has little or no ability in explaining the cross-sectional mean returns and that firm size (SMB) and book-to-market equity (HML) effects seem to describe it in a meaningful manner.

To evaluate the goodness of fit of this models, the cross-sectional R^2 measure employed by Jagannathan and Wang [17] and Lettau and Ludvigson [20] is used. The R^2 measures the fraction of cross-sectional variation in average returns that is explained by the model. In the discussed case, $R^2 = 65.1\%$ (this is close to the result ($R^2 = 71\%$) obtained for the US data [24]).

Figure 3 presents adjustment between the observed means of monthly returns (y -axis) and the expected monthly returns (x -axis) predicted by the model.

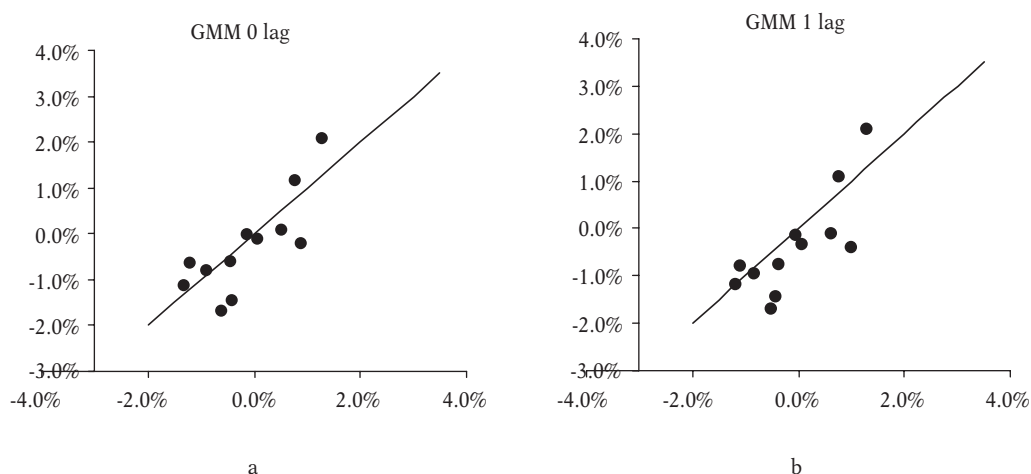


Figure 3. Average Monthly Returns (y-axis) $\times$ Predicted Monthly Returns (x-axis)

4. CONCLUDING REMARKS

The empirical study was carried out on data from December 2002 to January 2010. To account for heteroscedasticity and autocorrelation the unknown parameters of model were estimated using the GMM method. The results obtained from the study confirm the hypothesis that the Fama-French SMB and HML factors quite satisfactorily describe changes of portfolios returns in the sample period. The rate of return depends on the risk connected with the investments in stocks of small firms and in stocks of firms underpriced by the market with high BE/ME ratio. The risk premiums are significant in the case of the SMB and HML factors. The GMM point estimator indicates a premium close to 2% in the case of the HML and close to 0.7% in the case of the SMB. The risk premium connected with the market factor seems insignificant, however it cannot be excluded. The goodness of fit of the three-factor model evaluated by the cross-sectional R^2 measure is greater than sixty five percent.

Presented results should be considered as an introduction to a wider study. Investigations on whether portfolio constructions have an influence on the estimation results and the discussion on the choice of proper statistical tools will be the subject of the future work.

AGH University of Science and Technology, Faculty of Management

REFERENCES

- [1] Andrews D.W.K., [1991], *Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimation*, „Econometrica” 59, 817-858.
- [2] Banz R.W., [1981], *The Relationship Between Return and the Market Value of Common Stocks*, „Journal of Financial and Quantitative Analysis”, 14, 421-441.
- [3] Basu S., [1977], *Investment Performance of Common Stocks in Relation to their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis*, „The Journal of Finance” 32, 663-682.
- [4] Black F., Jensen M., Scholes M., [1972], *The Capital Asset Pricing Model: Some Empirical Tests*, in *Studies in the Theory of Capital Markets*, Michael C. Jensen, ed. New York: Praeger, pp. 79-121.
- [5] Blume M.E., Friend I., [1975], *The Asset Structure of Individual Portfolios with Some Implications for Utility Functions*, „Journal of Finance” 30, pp. 585-604.
- [6] Blume M.E., Stambaugh R.F., [1983], *Biases in Computed Returns: An Application to the Size Effect*, „Journal of Financial Economics” 12, 387-404.
- [7] Bolt T.W., Miłobędzki P., [2002], *Weryfikacja modelu CAPM dla giełdy warszawskiej*, Prace Naukowe Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Nr 952, 89-95.
- [8] Fama E.F., MacBeth J., [1973], *Risk, Return and Equilibrium: Empirical Tests*, „Journal of Political Economy” 81, 607-636.
- [9] Fama E.F., French K.R., [1992], *The Cross-Section of Expected Stock Returns*, „Journal of Finance” 47(2), 427-465.
- [10] Fama E.F., French K.R., [1993], *Common Risk Factors in the Returns on Stocks and Bonds*, „Journal of Financial Economics” 33, 3-56.
- [11] Fama E.F., French K.R., [1996], *Multifactor Explanations of Asset Pricing Anomalies*, „Journal of Finance” 51(1), 55-84.
- [12] Fiszeder P., [2009], *Modele klasy GARCH w empirycznych badaniach finansowych*, Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, Toruń.
- [13] Gibbons M.R., Ross S.A., Shanken J., [1989], *A Test of the Efficiency of a Given Portfolio*, „Econometrica” 57(5), 1121-1152.
- [14] Hall A.R., [2005], *Generalized Method of Moments*, Oxford University Press, New York.
- [15] Hansen L.P., [1982], *Large Sample Properties of Generalized Method of Moments Estimators*, „Econometrica” 50(4), 1029-1054.
- [16] Harvey C., Kirby C., [1995], *Analytic Tests of Factor Pricing Models*, Working Paper (Duke University, Durham).
- [17] Jagannathan R., Wang Z., [1998], *An asymptotic theory for estimating beta-pricing models using cross-sectional regression*, „Journal of Finance” 57, 2337-2367.
- [18] Kowerski M., [????], *Trójczynnikowy model Fama i Frencha dla Giełdy Papierów Wartościowych w Warszawie*, „Przegląd Statystyczny” 55(4), 131-148.
- [19] Lakonishok J., Shapiro A.C., [1983], *Systematic risk, total risk and the size as determinants of stock market returns*, „Journal of Banking and Finance” 10, 115-132.
- [20] Lettau M., Ludvigsson S., [2001], *Consumption, aggregate wealth, and expected stock returns*, „Journal of Finance” 56, 815-850.
- [21] Lintner J., [1965], *The Valuation of Risky Assets and the Selection of Risky Investment in Stock Portfolios and Capital Budgets*, „Review of Economics and Statistics” 47, 13-37.
- [22] MacKinlay A.C., Richardson M.P., [1991], *Using Generalized Method of Moments to Test Mean-Variance Efficiency*, „Journal of Finance”, 46(2), 511-27.
- [23] Ogaki M., [1993], *Generalized Method of Moments: Econometric Applications*, Handbook of Statistics 11, 455-485.
- [24] Petkova R., [2006], *Do the Fama-French Factors Proxy for Innovations in Predictive Variables?*, The Journal of Finance, 2, 581-612.
- [25] Reinganum M.R., [1981], *Misspecification of Capital Asset Pricing: Empirical Anomalies Based on Earnings' Yields and Market Values*, „Journal of Financial Economics” 9, 19-46.
- [26] Rosenberg B., Reid K., Lanstein R., [1985], *Persuasive Evidence of Market Inefficiency*, „The Journal of Portfolio Management” 13, 9-17.

- [27] Shanken J., Zhou G., [2007], *Estimating and Testing Beta Pricing Models: Alternative Methods and their Performance in Simulations*, „Journal of Financial Economics” 84, 40-86.
- [28] Sharpe W.F., [1964], *Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk*, „Journal of Finance” 19, 425-442.
- [29] Schmidt P., [1976], *Econometrics*, Marcel Dekker, New York.
- [30] Urbański S., [2007], *Time-Cross-Section Factors of Rates of Return Changes on Warsaw Stock Exchange*, „Przegląd Statystyczny” 54(2), 94-121.
- [31] Zarzecki D., [1999], *Metody wyceny przedsiębiorstw. Zarys teorii a praktyka*, Eds.: Accountancy Development Foundation in Poland, Warsaw, ISBN: 83-7241-099-2.
- [32] Zellner A., [1962], *An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias*, „Journal of the American Statistical Association” 57, 348-368.

Praca wpłynęła do redakcji w marcu 2010 r.

MODEL CAPM ORAZ MODEL FAMY I FRENCHA NA WARSZAWSKIEJ GIEŁDZIE PAPIERÓW WARTOŚCIOWYCH

Streszczenie

Tematem prezentowanej pracy jest weryfikacja trójczynnikowego modelu Famy Frencha dla danych z Warszawskiej Giełdy Papierów Wartościowych. Okres badania obejmuje lata 2002-2010. Do estymacji nieznanymi parametrów modelu zastosowano uogólnioną metodę momentów (GMM). Przyjęto założenie istnienia heteroskedastyczności i autokorelacji szeregów czasowych biorących udział w badaniu. Ponadto dopuszczono możliwość istnienia korelacji czynników objaśniających z błędami losowymi występującymi w modelu regresji. Uzyskane wyniki potwierdziły tezę, że trójczynnikowy model Famy Frencha zadowalająco opisuje zmiany stóp zwrotu na rynku polskim w badanym okresie. Wynik tego badania należy jednak traktować jako wstęp do bardziej wnikliwych analiz.

Słowa kluczowe: trójczynnikowy model Famy-Frencha, Uogólniona Metoda Momentów, ryzyko systematyczne, premia za ryzyko

THE CAPM AND FAMA-FRENCH MODELS IN POLAND

Summary

The main objective of this paper is to verify the performance of the Fama-French model for the Polish market. The estimates for individual stock returns are obtained using the monthly data from the Warsaw Stock Exchange for the period December 2002 to January 2010. The Generalized Method of Moments is used to test hypotheses that lead to the validation of the Fama-French model. We find that the cross-sectional mean returns are explained by exposures to the three factors, and not by the market factor alone. These results are consistent with previous studies of developed markets.

Key words: Fama-French three-factor model, Generalized Method of Moments, risk premium

PIOTR TARKA

LATENT VARIABLE MODELS – ISSUES ON MEASUREMENT AND FINDING EXACT CONSTRUCTS IN CUSTOMERS' VALUES

1. INTRODUCTION TO MEASUREMENT MODEL AND LATENT CONSTRUCT

In some accounts, measurement can be defined as the assignment of numbers to categories of observations. The properties of the numbers become the properties of the measurement – nominal, ordinal, interval, ratio, and so on [17]. In measurement model the inference is to relate the scores to the construct (here otherwise called latent variable or latent construct). The measurement model must help in understanding and evaluation the scores that come from the item responses. In short they should tell us about the construct, and the use of the results in practical applications. Simply put, the measurement model must translate scored responses to locations on the construct (variable) map. Some examples of measurement models are:

- the “true – score” model of classical test theory,
- the “domain score” model,
- factor analysis models,
- item response models,
- latent class models.

Typically we have two types of models. The first approach focuses on the items and their relationship to the construct [5; 6]. The second approach focuses on the scores and their relationship to the construct. The intuitive foundation of the instrument-focused approach is what might be called simple score theory. There needs to be some sort of an aggregation of information across the items, but the means of aggregation is either left vague or assumed on the basis of historical precedent to be the summation of item scores. Simple score theory is more like a folk theory, but nevertheless exerts a powerful influence on intuitive interpretations.

The simple score theory approach was formalized by classical test theory (also known as true score theory). This approach was founded by Spearman [14; 15; 16]. He borrowed a perspective from the fledgling statistical approach of the time and posited that an observed total score on the instrument X , was composed of the sum of a “true score” T and an “error” E . The introduction of an error term allows for a quantification of inconsistency in observed scores, which is part of the solution to the problem with Guttman scales. Guttman scaling focused attention on the meaningfulness of the results from the instrument (i.e., its validity), whereas classical test theory models the statistical nature of the scores focused attention on the consistency

of the results from the instrument (i.e., its reliability). There has been a long history of attempts to reconcile these two approaches. One notable early approach is that of Thurstone [22], who clearly saw the need to have a measurement model that combined the virtues of both.

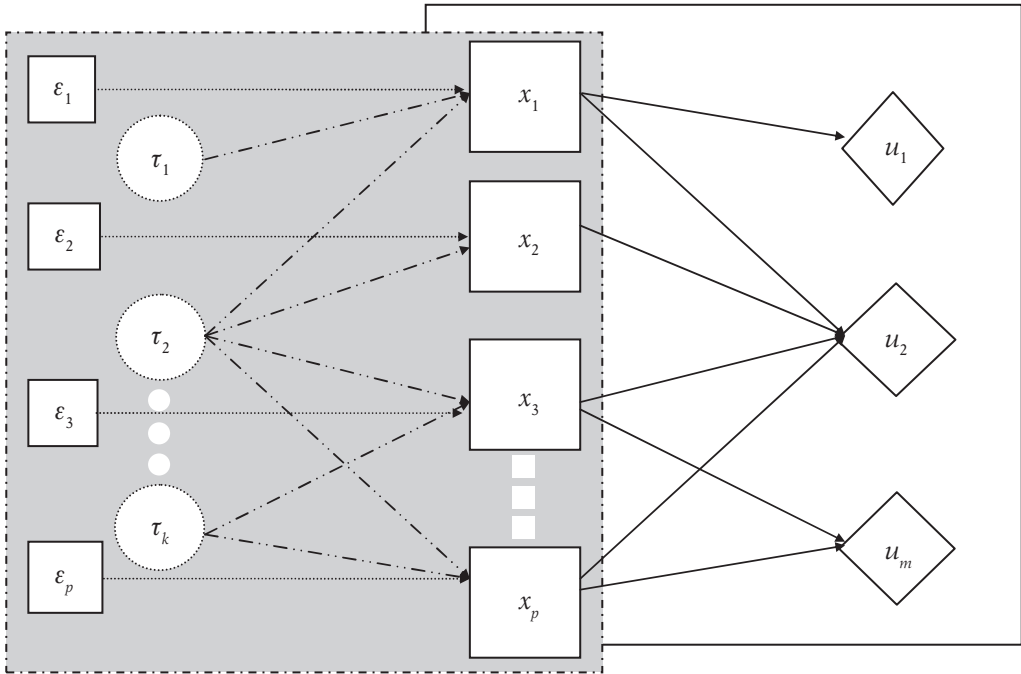


Figure 1. Measurement model and scale

Source: [21].

In statistically defined model, it must be assumed that x will be a vector of order p of the observed variables. The measurement model is therefore:

$$x = B\tau + \varepsilon, \quad (1)$$

where τ is a vector of order k of the true scores, B is a $p \times k$ pattern matrix which defines the relationship between x and τ , and ε is a vector of the measurement errors. The true scores and the measurement errors are not directly observable and must be estimated from the data. The assumptions are analogous to the classical test theory which was:

$$E(\tau) = \mu, E(\varepsilon) = 0, \text{cov}(\tau, \varepsilon) = 0, (\tau) = \Phi \text{ and } \text{cov}(\varepsilon) = \Psi. \quad (2)$$

Thus, the covariance structure of the observed variables is [15]:

$$\text{cov}(x) = E(B\tau + \varepsilon)(B\tau + \varepsilon)' = B\Phi B' + \Psi = \Sigma. \quad (3)$$

The covariance matrix Σ is assumed to be non-singular and positively definite:

$$\text{rank}(\Sigma) = p \text{ and } \Sigma > 0. \quad (4)$$

Unlike, the measurement scale (otherwise called latent construct) is a linear combination of the items. The collections of items into a composite score is intended to reveal levels of theoretical variables not readily observable by direct means (often referred to as scale). We develop scales when we want to measure phenomena that we believe to exist because of our theoretical understanding of this world, but that we can not assess directly. In general, we have m scales as a vector $u = A'x$, where A is a $p \times m$ matrix of the weights. The case of one scale is denoted by $u = a'x$ where a is a vector of the weights. It is important to distinguish between the concepts of the measurement scale from the measurement model. The model discriminates the underlying structure of the measurement from the use of the items [figure 1]. The scales weighted by the corresponding pattern elements of the model, denoted by $u = B'x$ are termed true score images. They are used to assess the structural validity of the measurement model.

2. ADJUSTMENTS IN THE MEASUREMENT MODEL DEPENDING ON DISTANCE BETWEEN RESPONDENT AND RESPONSE

There are basic assumptions as for the corrected model. First of all the measurement model must enable one firstly to interpret the distance between respondent and response on the construct (variable) map and secondly the measurement model must enable one to interpret distance between different responses on the construct map, and also the difference between different respondents. In order to find sense between these above two, we should have looked into what distance might mean in the context of a construct map.

On a construct map, distance between respondents and responses indicates the probability of making that response. To express this as an equation, we assume the probability of the response ($\text{Pr}(\text{response})$) is given by some function f of the difference between the respondent and the response [24]:

$$\text{Pr}(\text{response}) = f(q - d). \quad (5)$$

where interpretation for this formula can be given as follows:

1. zero distance between person and response would mean that person is likely to endorse statement with a certain probability,
2. respondent *above* response would indicate a greater probability,
3. respondent *below* response would indicate a lesser probability.

Hence, we can say that the model must have qualitative features from 1 to 3. However, these qualitative features are not sufficient to preserve the idea of a “map.” For that the requirements 2 and 3 would need to be more detailed, giving a specific metric form.

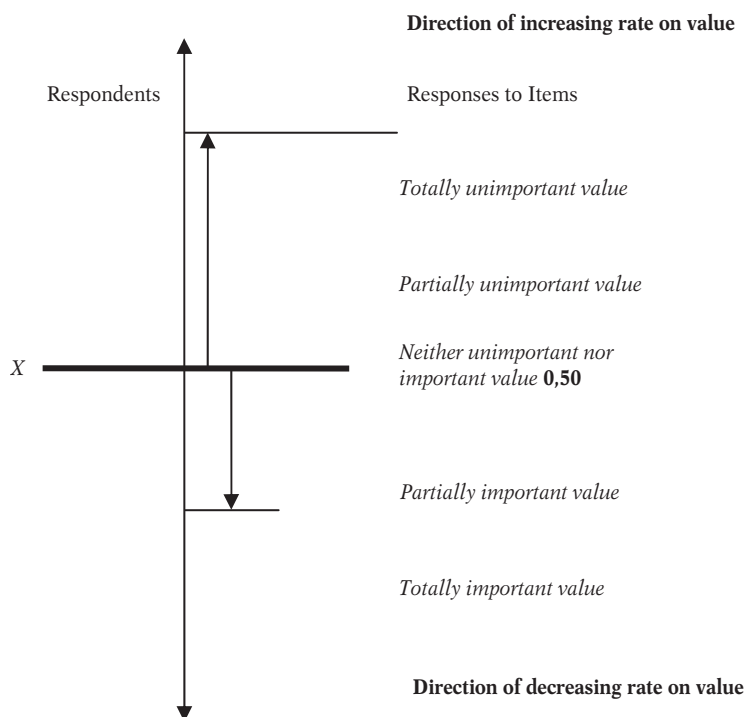


Figure 2. Different item locations with respect to a single person

Source: [24].

To consider distance between person and item, we can interpret it by the following figure 2. This figure illustrates the situation for one person approximately in the middle of the construct. For this person, items that are at a similar level would be expected to elicit agreement at about 0,50 probability, whereas items that are above would tend to result in a positive response with a greater probability and the opposite for those below.

We can also consider distances between item responses themselves. Figure 3 illustrates the distance between two item responses, considered by two persons. One specific qualitative consequence of this is that the order (on the map) of the item responses must remain the same for all respondents, and the order of the respondents (on the map) must remain the same for all item responses.

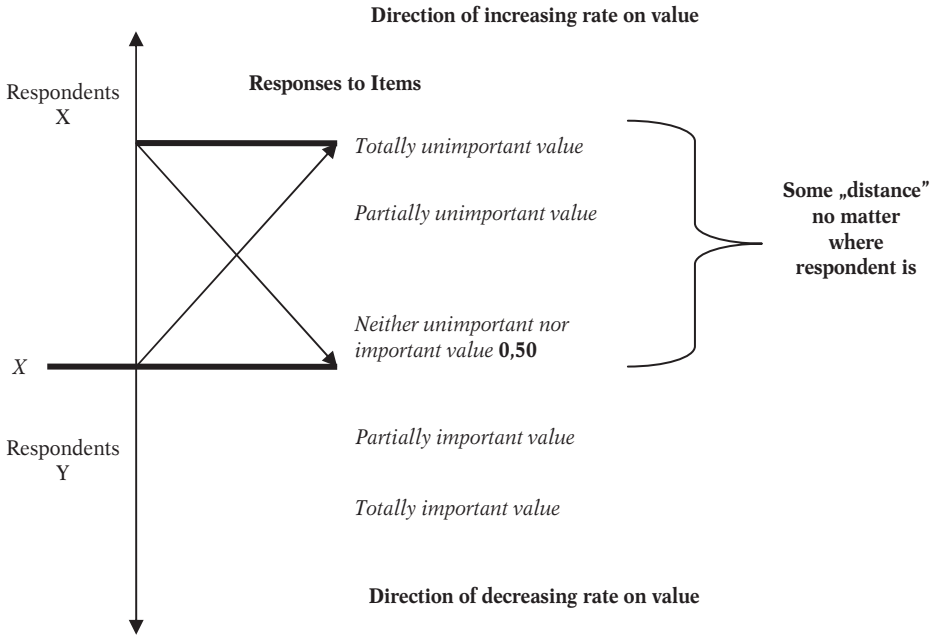


Figure 3. Different item locations with respect to two persons

3. FACTOR ANALYSIS IN THE DETECTION OF LATENT CONSTRUCT(S)

Factor analysis belongs to a group of latent variable models (table 1). The aim of factor analysis is to explain the outcome of p variables (items) in the data matrix using the so-called factors. Ideally all the information in data matrix can be reproduced by a smaller number of factors. These factors are interpreted as latent (unobserved) common characteristics of the observed variables. The case just described occurs when every observed $x = (x_1, \dots, x_p)$ can be written as:

$$x_j = \sum_{\ell=1}^k q_{j\ell} f_{\ell} + \mu_j, \quad j = 1, \dots, p \quad (6)$$

Here f_{ℓ} for $\ell = 1, \dots, k$ denotes the factors. The number of factors, k should always be much smaller than p . It is then possible to create a representation of the observations that is similar to the one in equation (6) by means of principal components, but only if the last $p - k$ eigenvalues corresponding to the covariance matrix are equal to zero. Considering a p - dimensional random vector X with mean μ and covariance matrix $\text{Var}(X) = \Sigma$ enables to form a model similar to (6) being written for X in matrix notation, namely

$$X = QF + \mu, \quad (7)$$

where F is the k – dimensional vector of the k – factors. When using the factor model (7) it is often assumed that the factors F are centered, uncorrelated and standardized: $E(F) = 0$. The existing factor model is defined by covariance matrix of model:

$$\Sigma = E(X - \mu)(X - \mu)^T = QE(FF^T)Q^T = QQ^T = \sum_{j=1}^k \lambda_j \gamma_j \gamma_j^T. \quad (8)$$

Table 1

Classification of latent variable models

	Observed variables x	
Latent variables f	Metrical (interval/ratio)	Categorical (nominal/ordinal)
Metrical (interval/ratio)	Factor analysis	Latent trait analysis
Categorical (nominal/ordinal)	Latent profile analysis	Latent class analysis

Source: [2; 3; 9; 10; 13].

It is also common in factor analysis to split the influences of the factors into: common and specific ones. There are, for example, highly informative factors that are common to all of the components of X and factors that are specific to certain components. The factor analysis model used in praxis is a generalization of (7):

$$X = QF + U + \mu \quad (9)$$

where:

Q – $(p \times k)$ matrix of the (non – random) loadings of the common factors $F(k \times 1)$,
 U – matrix of the random specific factors.

It is also assumed that the factor variables F are uncorrelated random vectors and that the specific factors are uncorrelated and have zero covariance with the common factors. More precisely, it is assumed that:

$$\begin{aligned} E(F) &= 0, \\ \text{Var}(F) &= \tau_k, \\ E(U) &= 0, \\ \text{Cov}(U_i, U_j) &= 0, \\ \text{Cov}(F, U) &= 0. \end{aligned} \quad (10)$$

The generalized factor model (7) together with the assumptions given in (8) can be constituted:

$$\underset{(p \times 1)}{X} = \underset{(p \times k)}{Q} \underset{(k \times 1)}{F} + \underset{(p \times 1)}{U} + \underset{(p \times 1)}{\mu} \quad (11)$$

where:

μ_j – mean of variable j ,

U_j – j -th specific factor,

F_1 – ℓ -th common factor,

$q_{j\ell}$ – loading of the j -th variable on the ℓ -th factor.

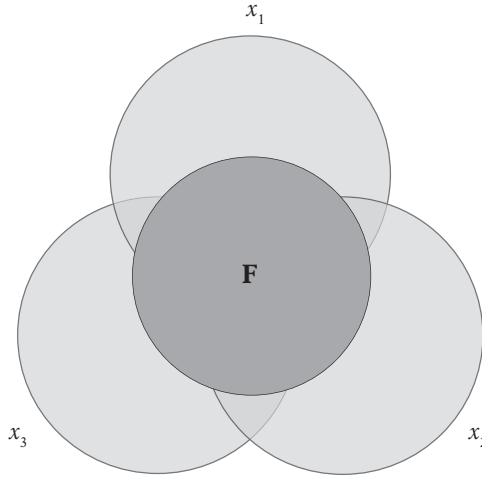


Figure 4. Depiction of the commonality among three measures of a single construct

The central concern of factor analysis in the process of modeling factors, is sometimes referred to as latent variables. Factors are influences that are not directly measured but account for commonality among a set of measurements. In figure 4 the commonality among measures of a construct is depicted in two ways. On the left, a diagram is used to illustrate a pattern of shared variance among scores on the three measures. Each circle represents the variance in one of the measures, x_1 , x_2 and x_3 . The overlap of the circles represents shared variance, or covariance. The shaded area, labeled “F”, represents the area of overlap involving all three circles. It is this area that corresponds to a factor or otherwise called “latent construct”.

4. EFA AND CFA AS LATENT METRICAL VARIABLE MODELS

There are two latent models based on metrical characteristics: CFA (Confirmatory Factor Analysis) and EFA (Exploratory Factor Analysis). Confirmatory factor analysis (CFA), otherwise referred to as restricted factor analysis, structural factor analysis, or the measurement model, typically is used in a deductive mode to test hypotheses regarding unmeasured sources of variability responsible for the commonality among a set of scores. It can be contrasted with exploratory factor analysis (EFA) which addresses the same basic question but in an inductive, or discovery-oriented, mode.

In order to illustrate associations among the measured variables making up complementary latent constructs a patch diagrams for CFA and EFA should be drawn. In the path diagram, the rectangles represent the measured variables, generically referred to as indicators. Ellipses represent unmeasured variables. The large ellipse represents a factor, whereas the smaller ellipses represent errors of measurement or uniqueness, which are unobserved sources of influence unique to each indicator. The single-headed arrows suggest causal influence, indicating specifically that each indicator is caused by two unmeasured influences:

- a causal influence it shares with the other indicators,
- an additional causal influence not shared with the remaining indicators.

The two-head curved arrows associated with the latent variables represent variances.

The path diagram can be translated directly into statistical form through a set of measurement equations, which specify the relations between factors and indicators. Consider the single equation, expressed in Bentler-Weeks [3] notation:

$$x_i = *F_1 + \dots + *F_j + u_i. \quad (12)$$

According to this equation, variability in the i th indicator is an additive function of j differentially weighted factors and the i th unique factor.

$$\begin{aligned} x_1 &= *F_1 + *F_2 + u_1 \\ x_2 &= *F_1 + *F_2 + u_2 \\ x_3 &= *F_1 + *F_2 + u_3 \\ x_4 &= *F_1 + *F_2 + u_4 \\ x_5 &= *F_1 + *F_2 + u_5 \\ x_6 &= *F_1 + *F_2 + u_6 \end{aligned} \quad (13)$$

Careful consideration of this equation illustrates a basic difference between CFA and EFA. In EFA, no restrictions are placed on the pattern of weights, denoted by “*”, associated with the factors. As such, each indicator is, to some degree, influenced by each of the j factors and, in CFA language, is “free” to be estimated in the analysis. Thus, the EFA measurement equations for a two-factor model involving six indicators, displayed in figure 5.

In contrast, CFA requires the imposition of certain restrictions on the pattern of weights, or factor loadings, a seemingly limiting characteristic that, in fact, pays substantial inferential dividends. Below equations are the CFA measurement equations for a prototypic two-factor CFA model involving six indicators, shown also in path diagram form in figure 5.

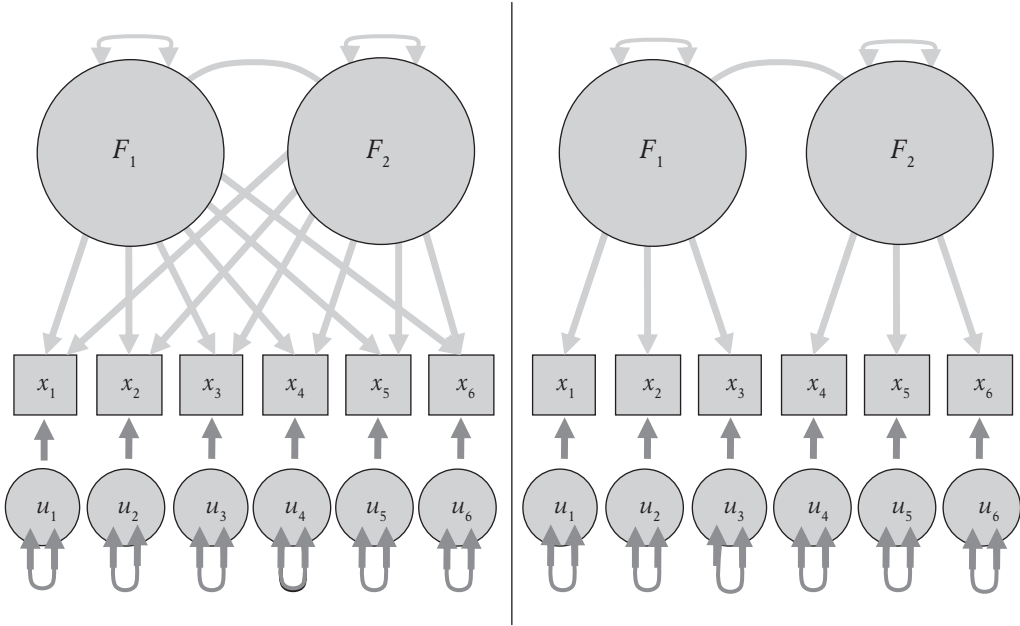


Figure 5. Path diagrams of two correlated factors as modeled using exploratory factor analysis (EFA) (left) and confirmatory factor analysis (CFA) (right).

$$\begin{aligned}
 x_1 &= *0F_1 + *0F_2 + u_1 \\
 x_2 &= *0F_1 + *0F_2 + u_2 \\
 x_3 &= *0F_1 + *0F_2 + u_3 \\
 x_4 &= *0F_1 + *0F_2 + u_4 \\
 x_5 &= *0F_1 + *0F_2 + u_5 \\
 x_6 &= *0F_1 + *0F_2 + u_6
 \end{aligned} \tag{14}$$

The zeroes indicate no influence of the factor on the indicator (in CFA language these parameters are “fixed” at zero). For instance, as can be seen in equations x_1 and x_6 (CFA) or the path diagram on the right in figure F_1 influences x_1 but not x_4 and F_2 influences x_6 but not x_3 .

This general pattern is an extreme example of what is referred to in EFA terms as simple structure. In EFA simple structure is a target of inductively oriented extraction and rotation algorithms, in CFA simple structure typically is assumed or imposed on the pattern of factor loadings.

5. LATENT CONSTRUCTS AND PROJECTED OR ACTUAL CORRELATIONS

In order to find latent construct(s), one must find an approximation $\tilde{\Psi}$ of Ψ , the matrix of the specific variances, and then correct the correlation matrix of X , by $\tilde{\Psi}$. Using the matrix as a starting point, factor analysis examines the patterns of covariation represented by the correlation among items. Each item is then correlated with conceptual single latent construct – variable. The observed correlations between items can be recreated by appropriately multiplying the paths linking each pair of items via the factor.

To compute correlations efficiently between observed item responses and factor representing a latent construct(s), one can posit that the sum of all item responses is a reasonable numerical estimate of the one, all – encompassing latent construct that is assumed to account for interitem correlations. In essence, this overall sum is an estimate of the latent construct's "score". Because the actual scores for all items are presumed to be determined by one latent construct, a quantity combining information from all items (i.e. an overall sum) is a reasonable of that latent construct's numerical value. With values those assigned to those causal pathways, one then can compute projected interitem correlations based on this one – factor model. These model – derived correlations are projections of what the actual interitem correlations should be if the premise of only one underlying variable is correct. The legitimacy of the premise can be assessed by comparing the projected correlations to the actual correlations. This amounts to subtracting each projected correlation from the corresponding actual correlation based on the original data. A substantial discrepancy between actual and projected correlations indicates that a single – factor model is not adequate, that there is still some unaccounted – for covariation among the items.

The residual matrix is supportive in latent constructs extraction and it can be treated in the same way as the original correlation matrix was treated, extracting e.g. a second factor and another corresponding to new latent construct(s). Once again, correlations between the items and that second latent construct can be computed and, based on those correlations, a matrix of proposed correlations can be generated. Those proposed correlations represent the extent of correlation that should remain among items after the second factor has been taken into account. If the second factor captured all of the covariation left over after extracting the first factor, then these projected values should be comparable to the values that were in the residual matrix. If not, further factors may be needed to account for the remaining covariation not yet ascribed to a factor. This process can proceed, with each successive factor being extracted from the residual matrix that resulted from the preceding iteration, until a matrix is achieved that contains only acceptably small residual correlations.

Also partial correlations can be explained when the effects of other items are taken into account. If "true" factors exist in the data, the partial correlations should be small, because the items can be explained by the items loading on the factors. The one exception regarding high correlations as indicative of a poor correlation matrix occurs when two items are highly correlated and have substantially higher loadings than other items on that factor. Then, their partial correlation may be high because they are not explained to any great extent by the other items, but do explain each other.

6. EXAMPLE ON LATENT CONSTRUCTS IN FACTOR ANALYSIS AND APPLICATION TO MARKETING SEGMENTATION

In order to detect, test and describe customers attitudes towards their professed values there was conducted: Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). Afterwards, the extracted and tested factors were described with Multidimensional Scaling (MDS) method.

As far as the set of various items pertaining to human (customers) values is concerned in EFA and CFA, they all have been derived from earlier study and survey of young people (members of academic youth – students in age between: 20-24) [20]. Initially 22 items across various fields of human values interests were developed and introduced in questionnaire. They were as follows: 1 – “Freedom”, 2 – “Independence”, 3 – “Success”, 4 – “Enjoying life”, 5 – “Pleasure”, 6 – “Family”, 7 – “Honesty”, 8 – “Helpfulness”, 9 – “Forgiveness”, 10 – “Peace”, 11 – “Justice”, 12 – “Beauty”, 13 – “Harmony”, 14 – “Politeness”, 15 – “Honouring parents”, 16 – “Obedience”, 17 – “Wealth”, 18 – “Tradition”, 19 – “Moderation”, 20 – “Humbleness”, 21 – “Accepting my place in life” and 22 – “Devotion”. All items were measured on 5 point Likert scale.

Author fulfilled all conditions for the correctness of the above undertaken analysis e.g.: 1) variables have index meaning, 2) the measurement of all variables was carried out on Likert scale, 3) sample exceeded the minimum of 100 respondents ($n = 232$).

The extracted and confirmed factors in EFA and CFA, were later described by MDS method on selected variables such as: 1) gender, 2) four types of personality, 3) three typically performed activities by respondents (see table 2). These variables belonged to group of psychodemographics area that are based on AIOD approach – “Activities-Interests-Opinions–Demographics” [11].

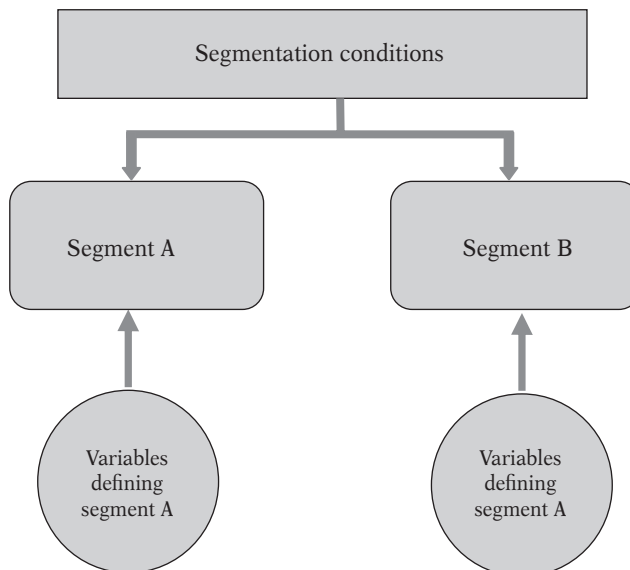


Figure 6. Relationship between segmentation conditions and its descriptors. Source: [11]

Thus one has described the relationship between descriptors and variables that constituted the principles of our segmentation [figure 6]. Besides in proceeding analysis there were only included those descriptors that might effectively differentiate segments. Descriptors that did not explain sufficiently differences in the process of segmentation of customers values were simply excluded from MDS.

Moreover there were used only values with highest factor loadings and all calculations were conducted in software: Statistica 9.0., SPSS 18.0 and Lisrel 8.0.

Table 2

Selected customers' variables from AIOD considered in segmentation analysis

Personality type	Performed activities	Gender
1. Technologically up-to-date: Aware of the possibilities that modern technology offers, and completely „on line” with the latest developments in technology	• Party: Going out – partying and having fun	Female
2. Extrovert: Friendly and outgoing; energized by being around other people; a person who is active and expressive	• Watching TV	Male
3. Advertising-conscious: Scan the print and electronic media for specials and bargains	• Reading book	
4. Locally oriented: Read local news and prefer to get it from newspapers rather than from television		
Measured on: Likert scale		Nominal scale

Source: [20].

Stage I – Exploratory Factor Analysis (EFA)

Before analyzing structure of constructs, a scree plot with eigenvalues was initially examined, where one might have concluded there were at least 4 of them. On the observed plot of eigenvalues, what can be seen is an elbow that is visibly cracked and rapid. That's why in the first row we thought the sample should be extracted with only four constructs consisting of 22 items. The rest constructs on the shallow slope, contribute very little to our model. The last big drop actually occurs between the fourth and fifth component, so using the first four ones is an easy choice. There is also a visible one prevailing construct above the others. This construct (with number one) explains much part among all eigenvalues.

Also comparing a visual presentation (a scree plot) to extracted constructs by **method of principal components** (PC) and its variance being explained, it is right to say again, that there are at least 4 retained constructs with eigenvalues greater than 1. Those constructs with eigenvalues lower than 1, were eliminated from our further analysis. As we can observe (when 22 items were considered at first) the greatest distance

(table 3) appears between values $\lambda_1 = 6,32$ and $\lambda_4 = 1,15$ and later between $\lambda_1 = 6,32$ and $\lambda_3 = 1,55$ respectively. Distance created between 1 and 4 is the greatest.

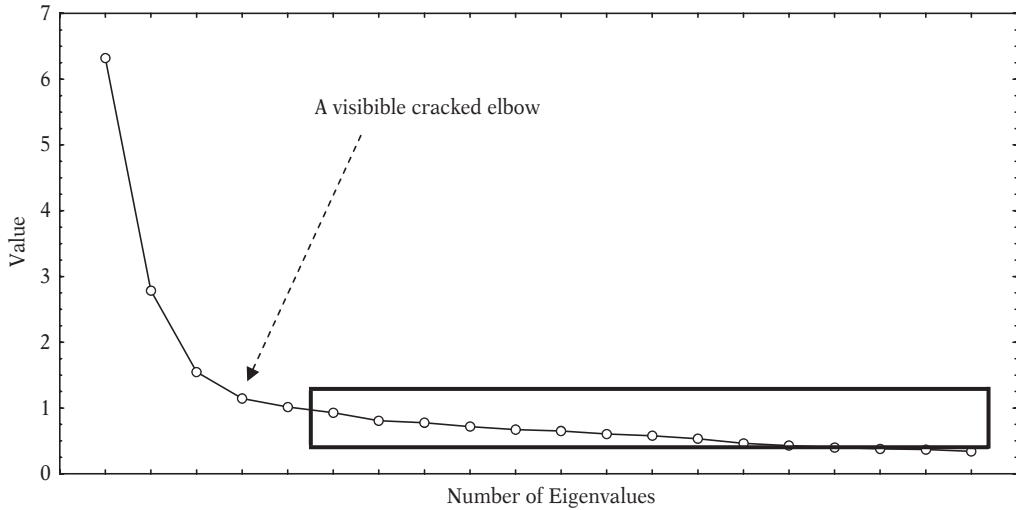


Figure 7. Plot of eigenvalues including all 22 item values

However after removing 8 items (observable variables), eigenvalues structure also changed. As a result the total variance level (for 4 constructs) was not accounted for (as before with 53,63%) but with 62,83% of variance.

Table 3

Eigenvalues extraction by principal components

	Eigenvalues at 22 items		
	Eigenvalue	% Total variance	Cum. % Total variance
λ_1	6,32	28,73	28,73
λ_2	2,79	12,66	41,39
λ_3	1,55	7,04	48,42
λ_4	1,15	5,21	53,63
Eigenvalues at 14 items – after reduction 8 other items			
λ_1	3,84	27,39	27,39
λ_2	2,45	17,49	44,88
λ_3	1,38	9,86	54,74
λ_4	1,13	8,08	62,83

Also the Kaiser's *KMO* criterion [7] enabled us to infer about higher level adequacy of above selected items for our model with 14 items. The index of *KMO* was:

$$KMO = \frac{\sum_p \sum_{h \neq p} r_{ph}^2}{\sum_p \sum_{h \neq p} r_{ph}^2 + \sum_p \sum_{h \neq p} \hat{r}_{ph}^2} = 0,74_{with\ 22\ items} \text{ and respectively } KMO = 0,79_{with\ 14\ items} \quad (15)$$

Here we can observe by the Kaiser's index, that it has increased when 8 items were excluded from further analysis. The level of *KMO'* interpretation pertaining to model quality assessment should be explained according to structure correlation matrix. They are: 0,9 – excellent, 0,8 – recommendable, 0,7 – decent, 0,6 – average, 0,5 – poor.

As a result we have focused only on 14 items (instead of 22) that were considered in further part of analysis (CFA).

In order to estimate correctly factor loadings in comparison with their items we've used four methods:

1. Principal Components (PC),
2. Maximum Likelihood (ML),
3. MINimum RESiduals (MINRES).
4. Generalized Least Squares (GLS).

As we can observe from table 4, factor loadings are slightly lower than in Principal Components method. Obtained values (when using ML, MINRES and GLS method) are lower than in PC. It is because factor analysis explains only communalities, not total variance. Hence the structure of final composite factors is changed and their interpretation must be also modified.

At last in the course of EFA (subjective analysis) we've identified initially four factors but in the next part of analysis CFA, two of them were correlated, which resulted in two final factors (table 7).

Table 4

Factor loadings structure after Varimax (PC, ML, MINRES) and Promax (GLS) rotation
(with marked loadings that are >0,50)

Items	PC Principal Components method				ML Maximum Likelihood method			
	$\hat{q}_1$	$\hat{q}_2$	$\hat{q}_3$	$\hat{q}_4$	$\hat{q}_1$	$\hat{q}_2$	$\hat{q}_3$	$\hat{q}_4$
1	,05	,18	,18	,84	,05	,27	,54	,20
2	,02	,22	,10	,85	,02	,23	,97	,08
3	,24	,70	-,07	,13	,20	,53	,13	,00
4	-,17	,74	,25	,18	-,11	,74	,14	,15
5	-,09	,74	,24	,17	-,03	,72	,14	,14

continued Table 4

Items	PC Principal Components method				ML Maximum Likelihood method			
	$\hat{q}_1$	$\hat{q}_2$	$\hat{q}_3$	$\hat{q}_4$	$\hat{q}_1$	$\hat{q}_2$	$\hat{q}_3$	$\hat{q}_4$
6	,30	,01	,74	,12	,36	,03	,12	,51
7	,20	,08	,81	,22	,23	,10	,17	,89
8	,64	,23	,24	-,02	,59	,19	,00	,18
9	,19	,64	-,25	,04	,13	,51	,06	-,08
10	,71	,09	,00	,12	,60	,09	,03	,08
11	,69	,22	,19	-,17	,67	,18	-,06	,05
12	,62	-,02	,37	-,09	,62	-,01	,04	,17
13	,74	-,02	,11	,06	,66	,01	,07	,11
14	,74	-,16	,02	,12	,64	-,09	,03	,05
Items	MINRES method				GLS method (Promax rotation)			
1	,05	,24	,61	,20	,10	,37	,61	,36
2	,02	,24	,85	,09	,06	,40	,93	,29
3	,21	,57	,12	-,02	,24	,57	,26	,13
4	-,12	,71	,17	,18	-,04	,77	,33	,24
5	-,04	,68	,17	,17	,04	,74	,33	,25
6	,35	,03	,14	,52	,42	,12	,21	,65
7	,24	,09	,19	,83	,34	,24	,30	,87
8	,59	,19	,02	,17	,62	,23	,09	,35
9	,13	,55	,06	-,10	,15	,52	,15	,00
10	,60	,10	,06	,06	,61	,13	,08	,25
11	,66	,18	-,08	,09	,69	,19	,01	,25
12	,61	-,02	,00	,21	,65	,03	,08	,36
13	,67	,00	,05	,09	,68	,06	,10	,28
14	,64	-,10	,05	,04	,63	-,05	,04	,22

Configuration of factor loadings after rotation (14 items): 1 – “Freedom”, 2 – “Independence”, 3 – “Success”, 4 – “Enjoying life”, 5 – “Pleasure”, 6 – “Beauty”, 7 – “Harmony”, 8 – “Obedience”, 9 – “Wealth”, 10 – “Tradition”, 11 – “Moderation”, 12 – “Humbleness”, 13 – “Acceptance”, 14 – “Devotion”

Stage II – Confirmatory Factor Analysis (CFA)

In this section we've implemented CFA to test the extracted factors. In CFA model we used:

1. Maximum Likelihood method (with Varimax rotation of Lawley's, where factor loadings proceeded through iterations). Thus we eliminated the uncertainty and subjectivity that strongly appeared in PC),
2. Generalized Least Squares (with Promax rotation) [1].

Applied **measures of goodness of fit** consisted of different types of indexes. For each of them results were displayed next to their mathematical notation. Their detailed description can be found in the work Sztemberg-Lewandowska [18].

The applied statistical test Chi-square (on the importance of fitting the model to data) in ML method (with function F_{ML}) reached the level $\chi^2 = 56,88$ at $df = 41$ and $p = 0,059$. It points to a value equal 0,059 of significance and is therefore greater than commonly accepted 0,05. One must therefore reject the hypothesis of no linear relationship between observed variables (items). It is noteworthy that this level (of 0,05) is exceeded only 0,009 [table 5]. Therefore it can be further inferred that correlations between factors are either weak or they are strong only for sake of two factors (as evidenced by the values of the correlation factors in table 6).

The value of significance (in GLS method with the function F_{GLS}) – reached a level of 0,152, and is much larger than 0,05.

Table 5

Goodness of fit measures for 4 factors

ML		
Chi-kwadrat	df	α
56,88	41	0,059
GLS		
50,28	41	0,152

Calculated indexes **GFI (Goodness-of-Fit)** and **AGFI (Adjusted Goodness-of-Fit)** by Joreskog and Sorbom) – (with the range [0, 1]) mark a good match to presented data model:

$$GFI = 1 - \frac{F[S, \Sigma(\hat{\theta})]}{F[S, \Sigma(\theta)]} = 0,93 \text{ and} \quad (16)$$

$$GFI = 1 - \left(\frac{c}{df_h}\right) \frac{F[S, \Sigma(\hat{\theta})]}{F[S, \Sigma(\theta)]} = 1 - \frac{c}{df_h} (1 - GFI) = 0,90$$

The high level of fit in the data is also proved by indexes such as **NFI (Normed Fit)** by Bentler and Bonett, – whose values typically are placed in range of [0, 1]

and *NNFI* – **Nonnormed Fit** Bentler and Bonett – where values fall even outside the range [0, 1]):

$$NFI = \frac{\chi_i^2 - \chi_h^2}{\chi_i^2} = \frac{F_i - F_h}{F_i} = 1 - \frac{F_h}{F_i} = 0,88 \quad (17)$$

$$NNFI = \frac{(\chi_i^2 / df_i) - (\chi_h^2 / df_h)}{\chi_i^2 / df_i - 1} = \frac{(F_i / df_i) - (F_h / df_h)}{(F_i / df_i) - (1 / (n - 1))} = 0,93 \quad (18)$$

Parsimony Goodness-of-Fit (by James, Mulaik and Brett) gives penalty for less restricted models (where more estimated parameters appear). In this index, values are always smaller than by indexes *GFI* and *AGFI*.

$$PGFI = \frac{df_h}{df_n} GFI = 0,69 \quad (19)$$

And finally the calculated **RMSEA (Root Mean Square Error of Approximation)** reached the level of 0,06. It measured the bad fit of data associated with model. Theoretically the closer its value to zero, the better the level of fit.

From table 6, we notice actually no correlations between other factors (which in fact proves to be orthogonal) except: 1) “Hedonistic Consumerism” and “Freedom – Independence” and 2) “Life Sensitiveness” and “Conservatism”. Therefore it was decided to merge factors with numbers: 1 and 3 and respectively 2 with 4 (table 7).

Table 6

Correlations between factors in ascending order

Factors	Level
Free Independence – Conservatism	0,13
Life Sensitiveness – Hedonistic Consumerism	0,10
Hedonistic Consumerism – Conservatism	0,11
Life Sensitiveness – Free Independence	0,12
Hedonistic Consumerism – Free Independence	0,52
Life Sensitiveness – Conservatism	0,54

Stage III – Interpretation of constructs in context of their utility for marketing activities

Initially in **EFA** model, we found four constructs altogether (table 7). Their contents were interpreted against the background of different areas aggregating particular

items. Construct one consisted of items such as: “Obedience”, “Tradition”, “Moderation”, “Humbleness”, “Acceptance”, and “Devotion”. Owing to specific structure of this composition and items characteristics, we’ve decided to give it a name “**Conservatism**”. But after **CFA** “Conservatism” construct was supplemented with two more items taken out from construct “Life sensitiveness”. Those added items were “Beauty” and “Harmony”. Respondents in this group highly valued attributes of nature and arts, serenity, peace, predictability and stabilization in every field of their live.

Table 7

Factors extracted in EFA and confirmed in CFA

Exploratory Factor Analysis			
Conservatism (1)	Hedonistic Consumerism (2)	Life Sensitiveness (3)	Freedom – Independence (4)
Obedience	Success	Beauty	Freedom
Tradition	Enjoying life	Harmony	Independence
Moderation	Pleasure		
Humbleness	Wealth		
Acceptance			
Devotion			
Confirmatory Factor Analysis			
Conservatism + Life Sensitiveness	Hedonistic Consumerism + Freedom – Independence		
Obedience	Success		
Tradition	Enjoying life		
Moderation	Pleasure		
Humbleness	Wealth		
Acceptance	<u>Freedom</u>		
Devotion	<u>Independence</u>		
Beauty			
Harmony			

Legend: Items underlined, mean they have been added on CFA model

Next construct – “**Hedonistic Consumerism**” was loaded with items such as: “Success”, “Enjoying life”, “Pleasure”, “Wealth”. This type of construct was also supplemented with items such as “Freedom”, “Independence”.

Stage IV – Multidimensional scaling (MDS) segmentation on constructs and AIOD variables

In MDS analysis, confirmed factors were analyzed across selected AIOD variables (table 2). The principles of MDS can be summarized in three-stage process:

- examination of structure containing objects or variables in matrix by defining their dimensions on the basis of similarities or preferences,
- visualization of data in n – dimensional space at $r < m$ relationships taking place between analyzed objects or variables,
- calculation distances $\hat{d}_{ij}$ in matrix and finding accurate level of data fit in the model.

In our example, we chose ALSCAL model (*Alternating Least Squares Scaling*) based on iterations. In this model one can take into account the measurement based on: 1) scale – both interval, ratio or order, 2) complete and incomplete data, 3) symmetrical and asymmetrical designated values of the data [12]:

$$S = \left\{ \frac{1}{m} \sum_{k=1}^m \left[\frac{\sum_i \sum_j (d_{ij,k}^2 - \hat{d}_{ij,k}^2)^2}{\sum_i \sum_j \hat{d}_{ij,k}^4} \right]^{\frac{1}{2}} \right\}. \quad (20)$$

ALSCAL is preceded by SSTRESS function [19]:

$$SS = \sum_i \sum_j \sum_k (d_{ij,k}^2 - \hat{d}_{ij,k}^2)^2. \quad (21)$$

More detailed information about the proces of model estimation, can be found in Zaborski's work [23].

In order to check the level of fit in our data with considered model we used additionally the squared correlation RSQ [12] which explains the proportion of variance of the scaled data in the partition that is accounted by its corresponding distance. In RSQ it is assumed that when using the method of MDS, an S – value less than 0,2 is “considered acceptable” for 2-dimensional solution. As a result RSQ square is high enough to capture substantial proportion of variance if it is above 85%. If two conditions are met (e.g. $S < 0,20$ and $RSQ > 0,85$ a 2-dimensional solution is a parsimonious one, that provides a sufficient explanatory power. A 3-dimensional one is “more complex” and adds little explanatory power. Thus a 2-dimensional solution (as it is in our case both for Male and Female group) can sufficiently reflect the information embedded in data.

Normally, ALSCAL is used to provide a visual representation of a complex set of relationships that can be scanned at a glance. Since maps on paper are 2-dimensional objects, this method translates technically to finding an optimal configuration of points in 2-dimensional space. However, the best possible configuration in 2 dimensions may be a very poor, highly distorted, representation of data. If so, this will be reflected in a high stress value. When this happens, one has two choices: one can either abandon

ALSCAL as a method of representing of data, or one can increase the number of dimensions. The problem is that with increasing dimensions, researcher must estimate an increasing number of parameters to obtain a decreasing improvement in stress. The result is model of the data that is nearly as complex as the data itself.

In our MDS analysis, the **male** group of respondents achieved $S = 0,163$ and $RSQ = 0,887$ in fourth iteration and respectively in **female** group $S = 182$ and $RSQ = 0,858$ in sixth iteration. Therefore we can infer that both models fit the data suitably. Obviously when the MDS map perfectly reproduces the input data, $f(x_{ij}) - d_{ij}$ that is for all i and j , the stress is then zero. Thus, the smaller the stress, the better the representation.

Table 8

Coordinates for items (observed variables) and selected AIOD variables

Items	Male		Female	
	Dim. 1	Dim. 2	Dim. 1	Dim. 2
1 – „Obedience”	0,54	–0,20	–0,53	–0,01
2 – „Tradition”	0,68	0,20	–0,86	0,15
3 – „Moderation”	0,62	–0,37	–0,84	–0,31
4 – „Humbleness”	0,96	–0,27	–0,85	0,45
5 – „Acceptance”	0,88	–0,24	–0,98	0,26
6 – „Devotion”	1,05	0,05	–1,12	0,02
7 – „Success”	–0,69	–0,17	–0,37	–0,56
8 – „Enjoyment”	–0,97	–0,45	0,33	–0,59
9 – „Pleasure”	–1,04	–0,31	–0,26	–0,44
10 – „Wealth”	–0,97	0,05	–0,54	–0,99
11 – „Beauty”	0,59	–0,60	–0,43	0,42
12 – „Harmony”	–0,03	–0,32	–0,41	0,70
13 – „Freedom”	–0,61	0,12	–0,01	0,53
14 – „Independence”	–0,70	–0,03	0,14	0,35
15 – „Technology”	0,50	1,36	0,93	0,39
16 – „Extrovert”	–0,49	1,14	0,96	0,79
17 – „Advert”	0,70	1,03	0,74	1,07
18 – „Local”	–1,03	0,81	1,11	–0,11
19 – „Party”	0,08	0,85	1,64	–0,05
20 – „Tv”	0,21	–1,22	0,02	–1,01
21 – „Book”	–0,25	–1,43	1,32	–1,05

Legend:

Item Values: 1 – “Obedience”; 2 – “Tradition”; 3 – “Moderation”; 4 – “Humbleness”; 5 – “Acceptance”; 6 – “Devotion”; 7 – “Success”; 8 – “Enjoyment”; 9 – “Pleasure”; 10 – “Wealth”; 11 – “Beauty”; 12 – “Harmony”; 13 – “Freedom”; 14 – “Independence”.

Personality Types: 15 – “Technology” (Technologically up-to-date); 16 – “Extrovert”; 17 – “Advert” (Advertising-conscious), 18 – “Local” (Locally oriented)

Performed activities: 19 – “Party” (going out for a party); 20 – “Tv” (watching Tv); 21 – “Book” (reading book)

When looking at a map that has non-zero stress, one must keep in mind that the distances among items are imperfect and representations of the relationships given by data are distorted. The greater the stress, the greater the distortion. However, one can rely on the larger distances as being also accurate. This is because the stress function sometimes accentuates discrepancies in the larger distances. Fortunately in our case, – the analyzed data and calculated Stress values for both groups reflect good fit.

Having based on 2 maps (with variables – see tables 2 and 7) one can make their appropriate interpretation. There is still one (technically speaking) important thing to realize about those maps. That is the orientation of the picture is arbitrary. Thus map representation of distances between customers' values and their personality characteristics as well their selected activities, must not be oriented such that **"Conservatism"** (with set of items) is up and **"Hedonistic Consumerism"** (set of items) is down. In fact, "Conservatism" might be down and "Hedonistic Consumerism" might be up. All that matters in this type of interpretation with these maps is which point is close to others and where "reasonable dimensions" or at least coherent configurations exist.

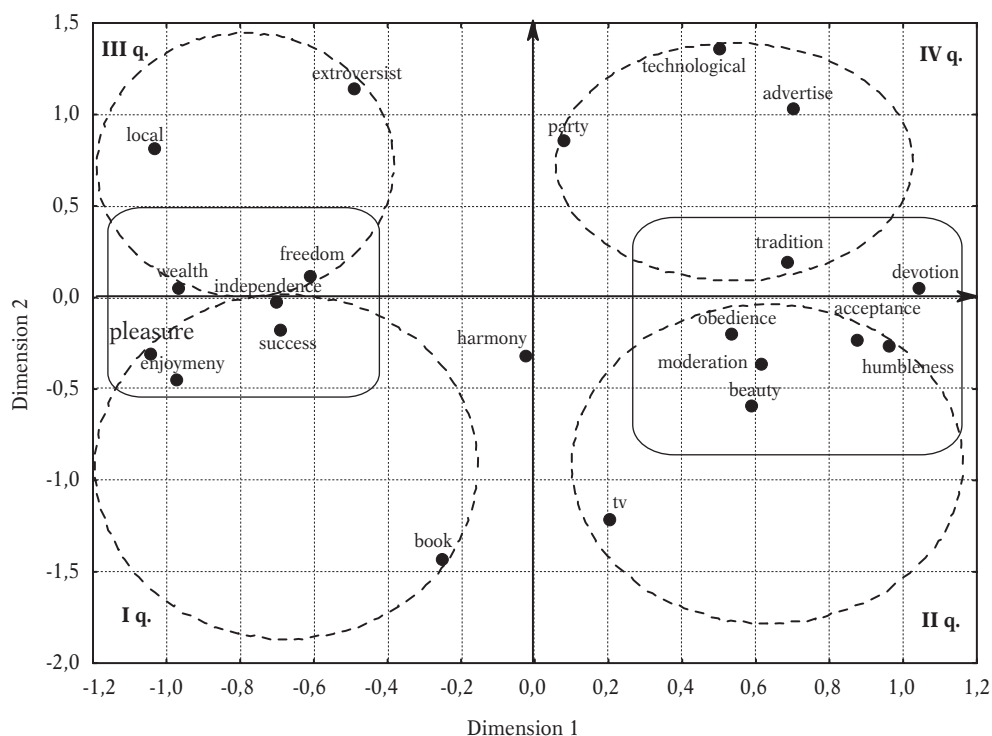
Stage V – Interpretation of MDS results in marketing context

In the map 1 and the dimension 1 (describing young men from **male group**), we may conclude that the those respondents valuing "Pleasure", "Enjoyment", "Success", "Freedom", "Wealth", "Independence" from one hand, are inherently "Extroverts" and "Locally oriented" and spend their spare time on reading books. Members of this group are generally well-read, hard – headed, goal-oriented and intelligent.

In turn, young people – e.g. in group of men who prefer "Tradition", "Acceptance", "Humbleness", "Moderation" and "Devotion" are more often "Advertising-conscious" and even "Technologically up-to-date", which makes them more susceptible to advertising campaigns and all kinds of technological product innovations such as new types of mobile phones. They prefer to spend their time on watching television or entertainment activities outside home.

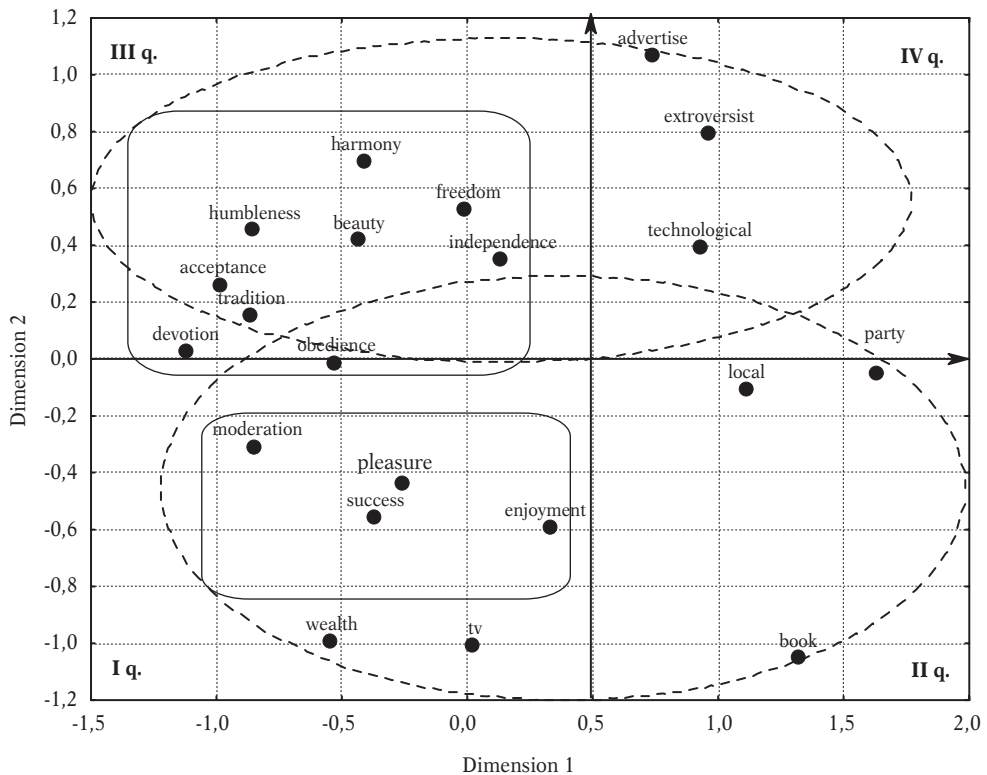
Men distinguish from young women by their professed values and their personality types as well their activities.

Moreover as we can observe (map 1) in first quarter, values such as: "Success", "Enjoyment", "Pleasure", or "Independence" are related to activity "Book reading". In second quarter there are values: "Obedience", "Acceptance", "Moderation", "Humbleness" and quite surprisingly value – "Beauty" that is connected with "Tv watching". In third quarter, "Independence", "Wealth" and "Freedom" is related to "Locally oriented" youth and "Extroverts".



Map 1. Male group of respondents

As far as **female group** of respondents is concerned (2 map), distinctions are not so clear as compared to male group. Young women seem to be more of complex character and personality type. Simply saying they have contrary configuration of traits as compared to men's group. First of all, they highly value: "Devotion", "Humbleness", "Tradition", "Acceptance", and "Harmony", but at the same time they're strongly interested in preserving in their life values such as: "Freedom", "Independence" or sustaining "Beauty". Obviously they strive to hold on to and mix "Conservatism" with "Hedonistic Consumerism" values. And those women (in the first quarter) who prefer "Pleasure", "Success" and "Enjoyment", also long for "Moderation", which might be a little bit confusing in men's map configuration.



Map 2. Female group of respondents

Actually values and AIOD variables configurations in women group, are mixing each other and they are not so constant as seen earlier in men's group. It also varies across the choice of selection within direction of analysis that is depending on particular dimension description. In short it changes from the perspective of visual interpretation by researcher. However it must be admitted that, women (in third and fourth quarters) are "Extroverts", "Technologically up-to-date", "Advertisement-conscious" and simultaneously hold on to "Conservatism" and "Hedonistic Consumerism" (jointly considered – as a certain way in their life style). Also young women in the first and second quarters are more focused on "Hedonistic Consumerism" and they are rather: "Locally oriented", "Book reading" and "Tv watching".

7. CONCLUSIONS

Finding the exact measurement model and method of estimation of the scores in constructs (otherwise called latent variables or latent constructs) can be extremely difficult challenge for any researcher. It is because of many various alternatives that

analyst must simultaneously consider and choose, according to which a final (appropriate or misleading) composition of constructs will be made. Certainly the measurement statistical model helps to some extent understand and evaluate the scores coming from the item responses making up a real construct. However by the time this construct will be generated, one must along the way overcome dilemmas about which type of model and method of estimation to choose, to make the best of it (e.g. either model focusing only on the items and their relationship to the construct or model focusing on the scores and their relationship to the construct).

Analytical problems also arise on the course of translation from scored responses to locations on the respondents' construct map – (which are made up of collections of items grouped into a composite score, not readily observable by direct means). Problems are usually related to:

- the distance between respondent and response on the construct map,
- distance between different responses on the construct map,
- difference between different respondents.

And number of p items (held in the data matrix) with quite various characteristics unfortunately don't simplify the identification of projected constructs and their later explanation. For instance using factor analysis, it is actually possible to locate latent constructs, but on the other hand it is sometimes difficult to judge whether selected for analysis items and grouped together by extracted construct are the right ones. It is because items can be (or should be) characterized not only on communality coefficients but also in relation to the interpretability of particular factors (constructs), in short, the items' played role (sense or meaning) in the interpretation of construct. Here once again the matter is open, whether choice of subjective or objective methods in factors (constructs) consideration would be a better solution. Usually these aspects are open and typically they are choices to be made.

As far as two constructs are concerned (confirmed in our analysis) one can assuredly make the following conclusions for marketing purposes. "Conservatism" and "Hedonistic Consumerism" constructs can be characterized towards certain targeted by companies segments of the market. Each construct represents not only a set of items, but above all it provides marketing information about suitable segment and its customers (with applications not only to youth segment but also to many others). Simply saying those constructs can work to marketing's advantages. Through them one may generate real ideas – concepts of marketing strategy and undertake effective actions presence of different segments in the area of: 1) promotion (e.g. expression of catching advertising slogans or selection of media communication such as: TV, press, radio or Internet); 2) choice of appropriate distribution channels based on indirect or direct customer preferred contact; 3) improved attributes in products and methods of product launch into the market or 4) setting the adjusted price level for products according to customers' personal preferences.

REFERENCES

- [1] Balicki A., [2009], *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne*, UG, Gdańsk.
- [2] Bartholomew D.J., Steele F., Moustaki I., Galbraith J.I., [2008], *Analysis of multivariate social science data*, 2nd ed. Chapman and Hall Books, New York.
- [3] Bentler P.M., Weeks D.G., [1980], *Linear structural equations with latent*, [in:] Long, J.S. (ed.), *Testing Structural Equation Models*, Sage Publications, New York.
- [4] Bąk A., [2007], *Application of latent variable models to consumers preferences analysis*, „Acta Universitatis Lodzensis: Folia Oeconomica”, Uniwersytet Łódzki, pp. 321-329.
- [5] Guttman L., [1945], *A basis for analyzing test-retest reliability*, „Psychometrika”, 10, pp. 255-282.
- [6] Guttman L., [1954], *Some necessary conditions for common-factor analysis*, „Psychometrika”, 19, pp. 149-161.
- [7] Kaiser H.F., [1960], *The varimax criterion for analytic rotation in factor analysis*, „Psychometrika” 23, pp. 187-200.
- [8] Kruskal J.B., [1964], *Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis*, „Psychometrika”, 29, pp. 1-28.
- [9] Loehlin J.C., [2004], *Latent variable models: an introduction to factor, path, and structural equation*, Lawrence Erlbaum Associates, New Jersey.
- [10] Marcoulides G.A., Moustaki I., [2002], *Latent variable and latent structure models*, Lawrence Erlbaum Associates, New York.
- [11] Sagan A., [2004], *Badania marketingowe – podstawowe kierunki*, UE Kraków.
- [12] Schiffman S.S., Reynolds M.L., Young F.W., [1981], *Introduction to multidimensional scaling: the theory, methods and applications*, Academic Press, New York.
- [13] Skrondal A., Rabe-Hesketh S., [2004], *Generalized latent variable modeling*, Chapman and Hall, New York.
- [14] Spearman C., [1904a], *The proof and measurement of association between two things*, „American Journal of Psychology”, 15, pp. 72-101.
- [15] Spearman C., [1904b], *General intelligence objectively determined and measured*, „American Journal of Psychology”, 15, pp. 201-293.
- [16] Spearman C., [1910], *Correlation calculated from faulty data*, „British Journal of Psychology”, 3, pp. 271-295.
- [17] Stevens S.S., [1946], *On the theory of scales of measurement*, „Science”, 103, pp. 667-680.
- [18] Sztemberg-Lewandowska M., [2008], *Analiza czynnikowa w badaniach marketingowych*, UE Wrocław.
- [19] Takane Y., Young F.W., de Leeuw J., [1978], *Nonmetric individual differences in multidimensional scaling: an alternating least squares method with optimal scaling features*, „Psychometrika”, 42, pp. 7-67.
- [20] Tarka P., [2008], *From ranking (Rokeach – RVS) to rating scales evaluation – some empirical observations on multidimensional scaling Polish and Dutch youth's values*, „Innovative Management Journal”, 1, 2, pp. 24-42.
- [21] Tarkkonen L., [1987], *On reliability of composite scales*, Statistical studies 7 ed. Finnish Statistical Society.
- [22] Thurstone L.L., [1947], *Multiple factor analysis*, The University Press, Chicago.
- [23] Zaborski A., [2001], *Skalowanie wielowymiarowe w badaniach marketingowych*, AE Wrocław.
- [24] Wilson M., [2005], *Constructing Measures: An Item Response Modeling Approach*, Lawrence Erlbaum Associates, New York.

Praca wpłynęła do redakcji w kwietniu 2010 r.

MODELE ZMIENNYCH UKRYTYCH – ROZWAŻANIA NAD POMIAREM I GENEROWANIEM KONSTRUKTÓW W SFERZE WARTOŚCI KONSUMENTÓW

Streszczenie

W artykule autor wymienia i opisuje różne typy modeli pomiarowych w zakresie tzw. zmiennych ukrytych. W pracy rozważa on także metodę i technikę analizy wyników w ramach identyfikacji ww. zmiennych (w literaturze występujących m.in. pod nazwą „konstruktów”). W niniejszej pracy omawiane są również zagadnienia związane z identyfikacją tego typu zmiennych z perspektywy poziomu pomiaru tj.: 1) odległości pomiędzy respondentami i ich odpowiedziami na tzw. mapie konstruktów (zmiennej); 2) odległości pomiędzy samymi odpowiedziami na mapie i 3) badaniu różnic pomiędzy respondentami. Dalsza część artykułu skoncentrowana jest na opisie dwóch klasycznych modeli analizy zmiennych ukrytych, opartych m.in. na metrycznych cechach pomiarowych: EFA (Eksploracyjnej Analizie Czynniskowej) i CFA (Konfirmacyjnej Analizie Czynniskowej). W pierwszej kolejności rozważana jest eksploracyjna analiza czynnikowa, którą to autor wykorzystuje do identyfikacji i opisu konstruktów (wartości konsumentów). W wyniku tej aplikacji (na podstawie 22 włączonych do analizy pozycji i ich wyników zgromadzonych z wcześniejszych przeprowadzonych badań empirycznych), wygenerowano wstępnie 4 konstrukty, którym nadano następujące nazwy: „Konserwatyzm”, „Wolność”, „Hedonistyczny konsumeryzm” i „Wrażliwość życiowa” – inaczej „Wrażliwość na otoczenie”. Następnie autor tworzy modeli CFA, gdzie redukuje liczbę zmiennych z 22 do 14 i ostatecznie tworzy 2 konstrukty: „Konserwatyzm” i „Hedonizm konsumpcyjny”. Na końcu artykułu, wyodrębnione konstrukty opisano na wybranych zmiennych z modelu AIOD, stosując do tego celu skalowanie wielowymiarowe. Konstrukty omówiono również w kontekście działalności marketingowej przedsiębiorstw.

Słowa kluczowe: Ukryte modele, konstrukty, wartości konsumentów

LATENT VARIABLE MODELS – ISSUES ON MEASUREMENT AND FINDING EXACT CONSTRUCTS IN CUSTOMERS' VALUES

Summary

In article author defines different measurement latent models and describes specify of measurement latent constructs. In literature some examples of these models are: the “true – score” model of classical test theory, the “domain score” model, item response model, factor analysis and latent class models. This work also presents method of estimation that should be undertaken in the identification process of latent constructs. Some aspects related with adjustments in the measurement model depending on distance between respondent and their responses, are also discussed. Author describes them from the prospect of 1) distance between respondent and response on the construct (variable) map; 2) distance between different responses on the construct map and 3) difference between different respondents. Next going on to further description, author considers two types of models based on metrical items characteristics: EFA and CFA. In the exploratory factor analysis as a key latent variable model in constructs detection and their formulation is defined where 4 latent constructs are extracted. These four detected constructs (based on earlier set of 22 value items) were given the following names: “Conservatism”, “Freedom-Independence”, “Hedonistic Consumerism”, and “Life Sensitiveness”. Secondly there is implemented CFA model which reduces number of value items from 22 to 14, containing only two latent constructs called “Conservatism” and “Hedonistic Consumerism”. Additionally those constructs were in the end, described with selected AIOD variables where MDS was applied. And at last constructs were defined in context of their utility for marketing activity.

Key words: Latent models, constructs, customers' values

DYDAKTYKA I NAUKA

MIROSŁAW SZREDER

LOSOWE I NIELOSOWE PRÓBY W BADANIACH STATYSTYCZNYCH

WPROWADZENIE

Jednym z istotnych etapów projektowania badania próbkowego (niewyczerpującego) jest określenie techniki wyboru próby (ang. *sampling technique*). Badacz musi najpierw rozstrzygnąć, czy będzie to któraś z grupy technik probabilistycznych (losowych) czy nieprobabilistycznych (nielosowych). O wyborze tym decydować mogą: zakres posiadanych lub dostępnych informacji o populacji, zamiar wykorzystania wnioskowania statystycznego do uogólnień na badaną populację, budżet przeznaczony na badanie, czas realizacji badania, i inne czynniki. Wskazanie konkretnej techniki albo kombinacji technik próbkowania, wiązać się będzie dalej z tak ważnymi kwestiami, jak: sposób określenia liczebności próby, możliwość włączenia do badania już posiadanych informacji (*a priori*) o populacji, wybór metod uzupełnienia (imputacji) braków w danych zebranych od respondentów.

Celem tego opracowania jest scharakteryzowanie wpływu, jaki na wybór techniki próbkowania mają coraz doskonalsze sposoby gromadzenia i przetwarzania danych o różnych populacjach poddawanych badaniom. Dynamiczny postęp technologiczny, którego doświadczamy od kilku dziesięcioleci, oddziałuje w sposób znaczący na wszystkie etapy procesu projektowania badania próbkowego, w tym na decyzję o sposobie doboru próby badawczej. Dzięki temu postępowi coraz więcej wiemy o wielu zbiorowościach poddawanych badaniu, i naturalna staje się potrzeba wykorzystania tej wiedzy już na etapie wyboru próby badawczej.

1. PRÓBA LOSOWA CZY NIELOSOWA?

Dla statystyka dylemat, czy próba powinna być losowa czy nielosowa, właściwie nie istnieje. Teoria klasycznego wnioskowania statystycznego oparta jest na modelu matematycznym, w którym zakłada się, że do próby dostają się jednostki wygenerowane przez mechanizm losujący, który każdej jednostce populacji daje taką samą szansę znalezienia się w próbie. Statystyka dostarcza znacznie więcej sformalizowanych narzędzi

wnioskowania w sytuacji, gdy do wspomnianego modelu matematycznego może się odwołać, aniżeli wtedy, gdy próba zostaje pobrana z populacji w inny sposób. Główną zaletą prób probabilistycznych, w szczególności próby losowej prostej, jest możliwość stosowania w dalszej analizie (wnioskowaniu) zasad rachunku prawdopodobieństwa (stąd nazwa: próba probabilistyczna, ang. *probability* – prawdopodobieństwo). Tam gdzie nie ma próby losowej, mechanizmu losującego, czy wreszcie zdarzeń losowych, tam nie ma zastosowania klasyczny rachunek prawdopodobieństwa. Jaka jest więc rola prawdopodobieństwa we wnioskowaniu statystycznym, i czy oznacza to, że próby inne niż losowe (nieprobabilistyczne) są mniej wartościowe w badaniach próbkowych? Zanim odpowiemy na te pytania i uzasadnimy, że próby nielosowe odgrywają także ważną rolę w praktyce wielu badań, przyjrzyjmy się najpierw roli prawdopodobieństwa we wnioskowaniu statystycznym.

2. PRAWDOPODOBIEŃSTWO I PRÓBY LOSOWE

Przed wszystkim warto zauważyć, że najważniejsze elementy wnioskowania statystycznego, w tym określenie wielkości próby, interpretacja wyników estymacji i testowania hipotez, charakterystyka wielkości błędu, wszystkie one w sposób bezpośredni odwołują się do prawdopodobieństwa. Oznacza to, że **u samych podstaw najważniejszych aspektów wnioskowania statystycznego leży założenie o możliwości stosowania rachunku prawdopodobieństwa** (założenie losowości próby badawczej). Prawdopodobieństwo otrzymuje tutaj interpretację częstościową – jedną z czterech najpopularniejszych interpretacji, obok: klasycznej, logicznej i personalistycznej (subiektywnej)¹. Interpretacja częstościowa oznacza, że prawdopodobieństwo zdarzenia definiowane jest jako granica częstości względnej zdarzeń elementarnych sprzyjających temu zdarzeniu, przy liczbie doświadczeń dążącej do nieskończoności. W praktyce rozumieć je należy jako częstość względną odpowiadającą dużej liczbie doświadczeń wykonanych w identycznych warunkach. W tym kontekście interpretuje się własności estymatorów wykorzystywanych we wnioskowaniu. Na przykład, nieobciążenie estymatora oznacza, że przy dużej (rosnącej do nieskończoności) liczbie losowań ustalonej wielkości próby, przeciętna wartość estymatora w tych próbach równa się wartości szacowanego parametru.

W estymacji przedziałowej prawdopodobieństwo występuje *explicite*, jako poziom ufności ($1 - \alpha$), który interpretować należy jako częstość względną liczby przedziałów ufności pokrywających szacowany parametr populacji w wielokrotnie powtarzanych losowaniach prób, z których każda określa własne granice przedziału ufności. Gdyby tę samą procedurę wielokrotnie powtarzanych prób zastosować do technik nielosowych próbkowania, w których mechanizm generowania obserwacji w próbie nie jest losowy, to nie byłoby żadnych podstaw do zastosowania rachunku prawdopodobieństwa. W konsekwencji, poziom ufności nie posiadałby interpretacji mówiącej cokolwiek o prawdopodobieństwie popełnienia (lub niepopołnienia) błędu w przedziałowej estymacji parametru. Dlatego zastosowanie podanej wyżej interpretacji do estymacji opartej na próbach nieprobabilistycznych (nielosowych) uznać należałoby za nieupoważnione.

¹ Szerzej na ten temat por. Szreder [1994] i [2010a].

Podobnie rzecz się ma z testowaniem hipotez statystycznych². Pojęcie prawdopodobieństwa występuje tu wprost jako poziom istotności (α) – prawdopodobieństwo popełnienia błędu polegającego na odrzuceniu hipotezy, gdy w rzeczywistości jest ona prawdziwa. Prawdopodobieństwo popełnienia tego błędu jesteśmy w stanie ustalić, bo potrafimy obliczyć częstość względną nietypowych prób w długim ciągu losowań. Nietypowych prób, czyli takich, których struktura – różna od struktury populacji – wskazuje na nieprawdziwość sprawdzanej hipotezy, podczas gdy faktycznie hipoteza ta jest prawdziwa. Wiemy z jaką częstotliwością takie nietypowe próby się pojawiają w losowym mechanizmie generowania obserwacji z populacji i na tej podstawie określamy prawdopodobieństwo błędu. Zwróćmy jednak uwagę, że prawdopodobieństwo α odnosi się wyłącznie do niedoskonałości mechanizmu losującego, odpowiedzialnego za tzw. błąd losowania (ang. *sampling error*). Gdyby ktoś zapytał, czy prawdopodobieństwo popełnienia błędu polegającego na odrzuceniu prawdziwej hipotezy uwzględnia też inne okoliczności, które do takiego błędu mogą prowadzić, jak na przykład: brak obserwacji na niektórych jednostkach próby (braki odpowiedzi respondentów), pomyłki respondentów, uchybienia w obliczeniach statystycznych, to oczywiście odpowiedź brzmi – nie. Błąd losowania i przypisane mu prawdopodobieństwo nie uwzględniają niczego poza niedoskonałością samego aktu losowania. A jeżeli tak, to zrozumiałe jest, że błąd ten nie występuje tam, gdzie w ogóle losowania nie ma. W technikach nieprobabilistycznych prawdopodobieństwo to (α) nie miałoby żadnej interpretacji. **Zdarzające się zastosowania teorii weryfikacji hipotez do prób nielosowych powodują, że traci się w tych warunkach możliwość określenia prawdopodobieństwa podjęcia błędnej decyzji, a także możliwość interpretacji przyjętego poziomu istotności.**

Do prawdopodobieństwa odwołujemy się także w ważnym zagadnieniu określenia niezbędnej wielkości próby badawczej. Zwróćmy najpierw uwagę, że pytanie: *Jak duża powinna być w danym badaniu próba?* jest niepełne i trudno jest na nie w ogóle odpowiedzieć, jeżeli nie poda się jakiegoś kryterium precyzji lub dokładności wnioskowania, które mają być spełnione. Dopiero w połączeniu z takim kryterium, pytaniu o liczebność próby można nadać odpowiedni sens logiczny i sformułować je za pomocą właściwych kategorii statystycznych. Na przykład: *Jaka powinna być liczebność próby, aby przeciętne odchylenie uzyskanych w badaniu ocen od prawdziwej wartości szacowanego parametru nie różniło się więcej niż o 2% lub o 5 jednostek miary, w której wyrażona jest badana cecha?* W badaniach próbkowych przyjęło się najczęściej stosować łącznie dwa kryteria:

- średni lub maksymalny błąd, rozumiany jako różnica między oceną z próby a prawdziwą wartością parametru w populacji (np. w badaniach opinii, w których szacowany jest wskaźnik struktury, przyjmuje się zwykle, że maksymalny błąd jest nie większy niż $\pm 3\%$);
- poziom ufności ($1 - \alpha$), czyli prawdopodobieństwo, z jakim przedział o postaci

ocena z próby $\pm$ błąd

zawiera prawdziwą wartość szacowanego parametru.

² Szerzej zagadnienie to zostało omówione w poprzednim numerze „Przeglądu Statystycznego”, M. Szreder [2010b].

Tak sformułowane zadanie określenia minimalnej liczebności próby, dla której spełnione są zadane kryteria, nie stanowi w większości schematów losowania poważnego problemu³.

Zwróćmy wszakże uwagę na istotny element tego wywodu. Założyliśmy *implicitę*, że badacz posługuje się próbą probabilistyczną, czyli że przy selekcji elementów do próby stosuje jedną z technik probabilistycznego wyboru. Bez tego założenia nie jest możliwa kontrola ani średniego błędu, o którym wspomnieliśmy, ani współczynnika ufności, który – jak stwierdziliśmy wcześniej – nie ma racji bytu we wnioskowaniu na podstawie prób nieprobabilistycznych (nielosowych). Oba kryteria, które wyżej wypunktowaliśmy, wiążą się bezpośrednio z losowaniem próby, a nie z jakimkolwiek innym sposobem jej wyboru.

3. PRZESŁANKI I KONSEKWENCJE NIELOSOWEGO WYBORU PRÓBY

Jak już wspomnieliśmy, rezygnacja z losowego generowania obserwacji do próby oznacza niemożność stosowania we wnioskowaniu pojęcia prawdopodobieństwa, przynajmniej w jego klasycznej i częstościowej interpretacji. W konsekwencji badacz, decydując się na nielosowy dobór próby, świadomie rezygnuje z rachunku prawdopodobieństwa, a co za tym idzie także z klasycznej Neymanowsko-Pearsonowskiej teorii wnioskowania statystycznego. Mogłoby się więc wydawać, że techniki nielosowego doboru próby są tym gorszym wariantem, wymuszonym okolicznościami, które nie pozwalają na zastosowanie technik probabilistycznych. Do niedawna było to dość powszechne rozumowanie, czego wyrazem był między innymi brak zainteresowania statystyków nielosowymi technikami próbkowania. W rzeczywistości okazuje się, że istnieją sytuacje wymuszające zastosowanie techniki nielosowej, ale **są też takie sytuacje, w których techniki te są szansą, a nie koniecznością**. Koniecznością stają się wówczas, gdy trudne lub niemożliwe jest zapewnienie każdej jednostce populacji równej możliwości dostania się do próby (np. z powodu braku operatu losowania, albo złej jego jakości). Niekiedy inne czynniki mogą decydować o tym wyborze, na przykład krótki czas na realizację badania albo skromne środki finansowe. Wszystko to może sprawić, że – na przykład – dla wewnętrznie zróżnicowanej populacji badacz zdecyduje się na zastosowanie techniki doboru kwotowego próby (nieprobabilistycznego), zamiast bardziej czasochłonnego losowania warstwowego. Szerszego wyjaśnienia wymagają zaś okoliczności, w których wybór techniki nielosowej nie jest niczym wymuszony, lecz jest świadomie stosowany, dzięki walorom danej techniki.

Analogicznie jak w punkcie 2, przyjrzyjmy się najpierw kwestii prawdopodobieństwa. Nielosowy sposób generowania zdarzeń – powtórzmy – nie upoważnia do stosowania klasycznej bądź częstościowej interpretacji prawdopodobieństwa. W praktyce jednak coraz rzadziej mamy do czynienia z „modelową” sytuacją, w której każde zdarzenie jest jednakowo możliwe. Mimo to jednak, pragniemy nadal posługiwać się pojęciem prawdopodobieństwa. Śledząc rozwój teorii prawdopodobieństwa nietrudno zauważyć, że zarówno filozofowie, jak i matematycy poszukiwali takiego sposobu

³ Por. Barnett [1991], Bracha [1998], Steczkowski [1995], Zasępa [1972].

pomiaru prawdopodobieństwa zdarzeń, który odzwierciedlałby bieżący stan wiedzy o określonych zdarzeniach⁴. Powszechnie było i jest dążenie do wykorzystania w ocenie prawdopodobieństwa całej wiedzy o zdarzeniu, nie tylko o jego prostym modelu (takim jak w interpretacji klasycznej) i o obserwowanej częstości względnej realizacji tego zdarzenia w niezmiennych okolicznościach (jak w interpretacji częstościowej). W ten sposób, w wyniku badań m.in. T. Bayesa, L.J. Savage'a, B. de Finettiego⁵ sformułowana została interpretacja personalistyczna prawdopodobieństwa, zwana inaczej subiektywną. Przez **subiektywne prawdopodobieństwo** (ang. *subjective probability*) tego, że jakiś sąd na temat zdarzenia A jest prawdziwy, rozumie się stopień pewności (ang. *degree of belief*) lub przekonania danej osoby o prawdziwości tego sądu. Zgodnie z tą interpretacją prawdopodobieństwo na temat zdarzenia A jest przypisane do danej osoby i może być różne dla różnych osób (ekspertów) w zależności od ich stanu wiedzy, doświadczenia, a nawet intuicji. W przypadku zdarzeń jednostkowych lub rzadko powtarzalnych jest to najczęściej stosowana interpretacja prawdopodobieństwa⁶. Przede wszystkim jednak jej popularność wiąże się z rosnącymi współcześnie zasobami informacji o różnych zdarzeniach, różnych populacjach, będących przedmiotem zainteresowania statystyków. Nieznane wcześniej możliwości gromadzenia, przetwarzania i przysyłania ogromnych zbiorów danych każą zwrócić uwagę na fakt, że coraz rzadziej badacz znajduje się w sytuacji zupełnego braku wiedzy o badanej populacji. Najczęściej wiedzę taką, mimo że częściową i niedoskonałą posiada lub może posiadać. I problemem nie jest to, czy ją wykorzystać, lecz jak ją wykorzystać. Jest to kluczowa sprawa dla zrozumienia zarówno coraz śmielszego odwoływania się statystyków do personalistycznej interpretacji prawdopodobieństwa, jak i coraz większej popularności prób nielosowych w badaniach niewyczerpujących. Ostatecznym bowiem **celem wnioskowania statystycznego nie jest osiągnięcie doskonałości w próbkowaniu, lecz jak najlepsze poznanie badanej populacji**. Gdyby mechanizm losowania próby był doskonały, to badacz nie miałby powodów, by w niego ingerować. A ingeruje coraz silniej, gdyż coraz bogatszą posiada wiedzę o populacji, użyteczną wiedzę, którą pragnie włączyć do próbkowania w celu poprawy jakości wnioskowania. Wiedzę tę ma prawo ująć także w formie probabilistycznej, stosując personalistyczną interpretację prawdopodobieństwa.

Ingerencja badacza w klasyczny schemat próby losowej prostej jest widoczna w powszechnie znanych schematach próbkowania, takich jak losowanie systematyczne, czy losowanie warstwowe. W tym ostatnim szczególnie dobrze widać przekonanie o niedoskonałości modelu próby losowej prostej w sytuacji, w której badacz zna (*a priori*) zróżnicowanie wewnętrzne populacji ze względu na cechy istotne dla celu badania.

⁴ G.W. Leibniz w 1955 r. pisał: *Ale to, co prawdopodobne sięga dalej; trzeba je wydobyć z natury rzeczy, a mniemanie osób o wielkim autorytecie jest jedną z rzeczy, które mogą się przyczynić do uprawdopodobnienia jakiegoś mniemania* (G.W. Leibniz [1955], s. 307). Również J. Bernoulli w swojej wcześniejszej pracy z 1713 r. pt. *Ars Conjectandi* („Sztuka przewidywania”) pisał o prawdopodobieństwie jako o stopniu zaufania do realizacji danego zdarzenia na podstawie posiadanej wiedzy o ogólnych okolicznościach tego zdarzenia.

⁵ Szerzej na ten temat, wraz z odpowiednimi odwołaniami bibliograficznymi por. M. Szreder [1994].

⁶ Przykłady zastosowań subiektywnej interpretacji prawdopodobieństwa w różnych zagadnieniach ekonomicznych podają m.in.: H. Kowalczyk [2010], A. Hołda i J. Pocięcha [2009], oraz M. Szreder [2004].

Ingerowanie zaś w mechanizm losowania, to nic innego jak stopniowe jego zastępowanie wyborem opartym na informacjach apriorycznych o badanej zbiorowości. Dlatego wraz ze zwiększaniem się zasobów informacji o badanych populacjach widoczna staje się tendencja do coraz częstszego korzystania w praktyce ze schematów losowania nieprostego⁷, a także z technik nielosowego doboru próby, takich jak wybór kwotowy. Gdy o wielokrotnie badanej populacji badacz posiada stosunkowo bogatą wiedzę, to kwotowy (nieprobabilistyczny) wybór próby może się okazać lepszy od schematów probabilistycznych.

Obawy przed stosowaniem technik nielosowego doboru próby dotyczą przede wszystkim kwestii możliwości oszacowania błędów pojawiających się przy uogólnianiu prawidłowości zaobserwowanych w próbie na całą populację. Techniki próbkowania nieprobabilistycznego nie zawierają tego składnika błędu, który w dotychczasowym rozwoju statystyki został najlepiej poznany i opisany – błąd losowania (nazywanego też błędem losowym). A ten właśnie błąd jest nie tylko podstawowym i zwykle jedynym błędem, który w sposób wymierny zostaje przypisany wynikowi wnioskowania w badaniach próbkowych⁸, ale jest ponadto podstawowym kryterium służącym określeniu liczebności próby (por. wyżej p. 2). W praktyce więc ośrodki badawcze stosujące nielosowe techniki próbkowania, jeżeli tylko doświadczenia ich są wystarczająco bogate, odwołują się do zgromadzonych doświadczeń i do wiedzy ekspertów, aby w sposób skwantyfikowany określić błąd, jaki może wynikać z posłużenia się w opisie populacji informacjami z próby (błąd próbkowania). Tak postępują znane polskie i zagraniczne ośrodki badawcze w odniesieniu do często stosowanego próbkowania kwotowego. Podkreślić jednak należy, że w każdym innym przypadku, gdy informacje o możliwym błędzie próbkowania są niewystarczające lub mało wiarygodne, charakterystykę populacji ograniczyć należy do opisu statystycznego, a nie wnioskowania statystycznego. Zaobserwowane prawidłowości w próbie odnieść można nadal do zbiorowości, której próba ta jest reprezentantem, jednakowoż bez przypisywania im wielkości błędu lub prawdopodobieństw ich prawdziwości. Nieuzasadnione jest w tych okolicznościach posługiwanie się metodami wnioskowania statystycznego.

4. PODSUMOWANIE

Zwiększające się zasoby informacji na temat różnych zbiorowości poddawanych badaniom statystycznym sprawiają, że w badaniach próbkowych źródłem informacji nie jest już wyłącznie próba statystyczna, jak często bywało w przeszłości. Posiadana informacja *a priori* o badanej populacji staje się równie ważnym źródłem informacji dla statystyka. W tym upatrywać należy rosnącej popularności nielosowych technik

⁷ O wnioskowaniu dla prób nieprostych w wielu swoich pracach pisze prof. Cz. Domański wraz z zespołem, por. np. Cz. Domański i K. Pruska [2000].

⁸ Pozostałe składniki błędu całkowitego, jak wiadomo, o wiele trudniej poddają się pomiarowi. Ilustracją tego faktu może być twierdzenie niektórych znanych statystyków, iż błąd losowy jest nadmiernie badany (*sampling error is „over-researched”*). Sformułowanie to pojawiło się m.in. w artykule znanych statystyków Richarda Platka i Carla-Erika Särndala pt. *Can a statistician deliver?* opublikowanym w jęz. polskim wraz z dyskusją przez „Wiadomości Statystyczne” nr 4, 2001 r.

próbkiowania, a także innych niż klasyczna i częstościowa interpretacji prawdopodobieństwa. Rosnące możliwości gromadzenia i przetwarzania informacji statystycznych powodować będą – jak się wydaje – upowszechnianie się modelu wnioskowania opartego na łącznej wiedzy o badanej populacji: *a priori* i z próby. Wiedza wstępna (*a priori*) wykorzystywana będzie na wszystkich etapach badania próbkowego.

LITERATURA

- Barnett V., [1991], *Sample Survey. Principles and Methods*, E. Arnold.
- Bracha Cz., [1998], *Metoda reprezentacyjna w badaniu opinii publicznej i marketingu*, Wyd. Efekt.
- Domański Cz., Pruska K., [2000], *Nieklasyczne metody statystyczne*, PWE, Warszawa.
- Hołda A., Pociecha J., [2009], *Probabilistyczne metody badania sprawozdań finansowych*, Wyd. Uniwersytetu Ekonomicznego w Krakowie, Kraków 2009.
- Kowalczyk H., [2010], *O eksperckich ocenach niepewności w ankietach makroekonomicznych*, „Bank i Kredyt”, nr 41, s. 101-122.
- Leibniz W.G., [1955], *Nowe rozważania dotyczące rozumu ludzkiego*, t. 2, PWN, Warszawa.
- Steczkowski J., [1995], *Metoda reprezentacyjna w badaniach ekonomiczno-społecznych*, PWN, Warszawa-Kraków.
- Szreder M., [1994], *Informacje a priori w klasycznej i bayesowskiej estymacji modeli regresji*, Wydawnictwo Uniwersytetu Gdańskiego.
- Szreder M., [2010a], *Metody i techniki sondażowych badań opinii*, PWE, Warszawa.
- Szreder M., [2010b], *O weryfikacji i falsyfikacji hipotez*, „Przegląd Statystyczny” nr 2-3, t. 57, s. 82-88.
- Szreder M., [2004], *Od klasycznej do częstościowej i personalistycznej interpretacji prawdopodobieństwa*, „Wiadomości Statystyczne” nr 8, s. 1-10.
- Zasępa R., [1972], *Metoda reprezentacyjna*, PWE, Warszawa.

GRZEGORZ KRZYKOWSKI

FILOZOFICZNE KONCEPCJE ESTYMACJI STATYSTYCZNEJ

Estymacja jako obszar analiz statystycznych danych pomiarowych jest wyznaczona trzema głównymi nurtami opisanymi w monografiach E.L. Lehmana [9], J.W. Tukey [15] i A. Gelmana ze wsp. [5]. Zagadnienia estymacji występujące w tych monografiach pojawiają się w wielu wcześniejszych pracach, ale pełne prezentacje kierunków estymacji zostały szczegółowo opisane przez wyżej wymienionych autorów.

1. REDUKCJA – TEORIA ESTYMACJI PUNKTOWEJ

Na wstępie postaramy się odpowiedzieć na pytanie, jakie główne koncepcje prowadzą do rozróżnienia podejścia do estymacji. Przyczyn rozróżnienia należy szukać w filozoficznej teorii nauki. W książce E.L. Lehmana [9] zaprezentowano teorię estymacji punktowej (porównaj również [2]). W tej koncepcji przyjmujemy, że rzeczywistości zjawisku badawczemu przypisujemy model statystyczny składający się z rodziny rozkładów. Zakładamy, że rodzina rozkładów jest indeksowana parametrem. Zwykle przyjmujemy techniczne założenie dotyczące własności tej parametryzacji, na przykład identyfikowalność parametryzacji (przykładowo porównaj [3]). Celem estymacji w tym podejściu jest wybór rozkładu (lub rozkładów) z przyjętej rodziny, zgodnie z różnymi kryteriami optymalizacji. Estymatorem jest rozkład lub zbiór rozkładów.

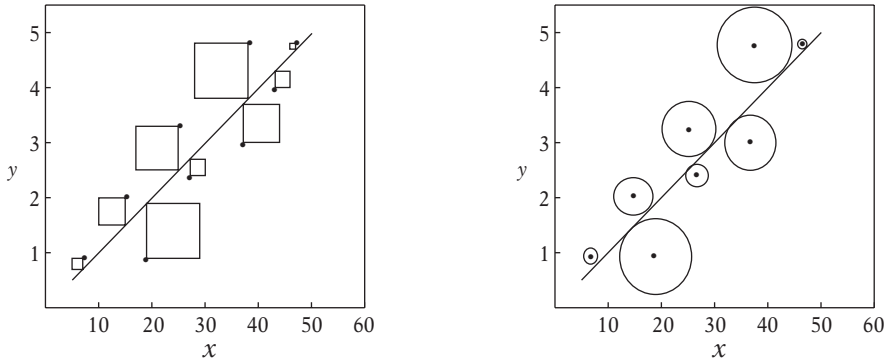
Wybranie rozkładu sprowadza się na ogół do wybrania parametru rozkładu. W najprostszym przykładzie przyjmujemy, że badanemu zjawisku eksperymentalnemu przypisujemy rodzinę rozkładów normalnych $\mathcal{P} = \mathcal{N}(\mu, \sigma^2)$, gdzie μ i σ^2 są parametrami identyfikującymi rozkład z tej rodziny ($\mu \in (-\infty, \infty)$, $\sigma^2 \in (0, \infty)$). Jeśli $\bar{x}$ stanowi realizację próby losowej prostej $(X_i)_{i=1}^n$ z tej rodziny, to uznajemy, że średnia z próby $\bar{x}$ identyfikuje grupę rozkładów normalnych o wartości oczekiwanej $\bar{x}$ i nieznaną wariancję.

W podejściu praktycznym, wartości estymatora nadajemy stosowną interpretację (jest to wartość oczekiwana), ale w sensie teoretycznym wybór estymatora w tym przypadku jest sprecyzowaniem podzbioru rozkładów w rodzinie rozkładów normalnych. Zwróćmy uwagę, że wynikiem estymacji jest podzbiór rodziny rozkładów.

W niektórych zagadnieniach jest to podzbiór jednoelementowy, w niektórych (jak w powyższym przykładzie) bardziej rozbudowany. Często staramy się nadać wynikowemu podzbirowi rozkładów pewną strukturę funkcyjną.

Na przykład przyjmujemy, że zbiór parametrów jest podzbiorem przestrzeni $\Theta \in \mathbb{R}^k$ i wynik estymacji traktujemy jako podzbiór rodziny rozkładów wyznaczony

przez podzbiór w Θ . Sytuacje takie, w szczególności gdy wyposażymy zbiór parametrów Θ w strukturę przestrzeni liniowej, nazywamy estymacją parametryczną. Jeśli nie zadajemy parametryzacji lub nie uwzględniamy jej struktury (na ogół liniowej), to estymację traktujemy jako nieparametryczną. Rozróżnienie estymacji parametrycznej i nieparametrycznej nie jest do końca jednoznaczne, jednak nie to jest głównym zagadnieniem w tym artykule. Przejdźmy do dalszych zagadnień.



Rysunek 1. Regresja średniokwadratowa i „średniokołowa”

Metod estymacji punktowej, czyli kryteriów optymalizacyjnych implikujących wybór rodziny rozkładów, jest bardzo dużo i każdy badacz indywidualnie decyduje o wyborze metody estymacji. Przytoczymy przykład z zakresu znajdowania linii regresji. Przyjmujemy, że pomiary (x, y) stanowią realizację próby losowej prostej $(X_i, Y_i)_{i=1}^n$ z rodziny rozkładów dwuwymiarowych oznaczonej $\mathcal{P}$, o której zakładamy tylko tyle, że wartości oczekiwane rozkładów brzegowych są skończone. Dla dowolnego rozkładu dwuwymiarowego z tej rodziny istnieją stałe $(a, b) \in \mathbb{R}^2$ takie, że $\mathbb{E}Y_i = a + b\mathbb{E}X_i$. Problem znalezienia oszacowania dla parametrów a i b jest przedmiotem estymacji punktowej. Zauważmy jednak, że w zagadnieniu tym rodzina rozkładów jest nieidentyfikowalna i *de facto* prowadzimy estymację nieparametryczną.

Wróćmy jednak do metody estymacji punktowej. Klasyczne podejście pochodzące od A.M. Legendrea¹ opiera się na metodzie optymalizacji średniokwadratowej, w której estymatory $\hat{a}$ i $\hat{b}$ są zadane klasycznymi wzorami Legendrea.

Pewne zdziwienie wywołałoby napisanie estymatorów dla współczynników a i b w postaci (patrz na przykład [11])

$$\hat{a} = \bar{y} - \hat{b}\bar{x}, \quad \hat{b} = \frac{-S_x S_y + \sqrt{(S_x S_y)^2 - 4S_x^2 S_y}}{S_{xx}^2}.$$

¹ Ostatecznie po dyskusjach historyków statystyki, za twórcę metody najmniejszych kwadratów uznano Legendrea, a nie jak wcześniej sądzono Gaussa [14]

Najczęściej recenzenci publikacji, w której zostałyby podana taka estymacja nie uznaliby jej za właściwą. Czy do końca mieliby rację?

Spójrzmy na rysunek 1. Okazuje się, że estymacja klasyczna (Legendrea) opiera się na minimalizacji średniokwadratowej odległości na osi Y, a przytoczony estymator parametrów a i b na minimalizacji średniokwadratowej odległości punktu od prostej (patrz na przykład [11]). Estymacja klasyczna jest zdecydowanie prostsza w obliczeniach analitycznych, ale można jej wiele zarzucić merytorycznie. Współcześnie, gdy praktycznie wszystkie obliczenia statystyczne są wykonywane w pakietach komputerowych, zastosować można bardziej skomplikowane obliczenia. Badacz powinien na tyle zapoznać się z zagadnieniem, nad którym pracuje, by móc określić jakie przesłanki leżą u podstaw tej czy innej metody estymacji punktowej. Podstawą wyboru określonej metody w estymacji punktowej powinno być merytoryczne uzasadnienie wyboru procedury optymalizacyjnej, a nie przywołanie klasycznych monografii, w których opisano konkretne standardowe metody estymacji punktowej.

W sensie filozofii nauki estymację punktową, polegającą na wyborze podzbioru rozkładów z zadanej rodziny rozkładów, należy uznać za formę rozumowania nazwanego redukcją. Redukcja jest metodą wnioskowania, która była analizowana i rozwijana przez polskich filozofów tworzących szkołę lwowsko-warszawską (patrz [16]).

W ogólnej metodologii nauk wyróżniamy pojęcia „racja” i „następstwo”. W obszarze znaczeniowym „racji” znajduje się rodzina rozkładów, a w obszarze znaczeniowym „następstwa” zespół danych pomiarowych. Według jednego z twórców szkoły, Jana Łukasiewicza, rozumowanie redukcyjne polega na znalezieniu „racji” na bazie „następstwa”. Rozumowanie to odpowiada procesowi estymacji punktowej².

2. INDUKCJA – EXPLORACJA DANYCH

O ile metody estymacji punktowej tworzone były przez dziesiątki, a nawet setki lat, o tyle estymacja, którą można nazwać eksploracją danych, powstała dość nagle. Powstanie metodologii EDA (Exploratory Data Analysis) wiążemy z pojawieniem się książki J.W. Tukeya o tym samym tytule³. Pojawienie się tego podręcznika uświadomiło statystykom podział estymacji na metodologię EDA i wcześniejszą CDA (Confirmatory Data Analysis). Przed publikacją monografii Tukeya [15] nie uważano za konieczne wyróżnienie tego podziału. Metoda eksploracji danych stosowana była prywatnie, niejako lokalnie, w etapie wstępnym poprzedzającym metodę estymacji punktowej. Takie wykorzystanie eksploracji danych wynika ze złożoności postępowania badawczego. Metoda EDA polega na analizie wyników pomiarów bez wchodzenia w naturę probabilistyczną pozyskanych danych badawczych. Można więc zapytać, co jest wynikiem takiej estymacji? W sensie teoretycznym odpowiedź nie jest oczywista, ale w sensie praktycznym owszem. Wynikiem estymacji są wszelkiego rodzaju pod-

² Konsekwencje koncepcji Łukasiewicza prowadziły do trudności w klasyfikacji nauk, co było przedmiotem dyskusji w ramach szkoły lwowsko-warszawskiej (patrz [1]).

³ Książka [15] powstaje po wielokrotnych wcześniejszych publikacjach stanowiących pierwotne składowe ostatecznej publikacji. Fragmenty te funkcjonowały w środowisku statystycznym w postaci niepublikowanych odbitek przez ponad 20 lat.

sumowujące wykresy. Należy podkreślić, że stosowanie eksploracji danych kończy się podsumowaniem w postaci opisu graficznego.

Wszelkie dalsze spostrzeżenia badawcze są już konsekwencją przyjrzenia się wykresom zagregowanych danych. Niektóre formy graficzne zaproponowane przez Tukeya nie utrzymały się w takiej postaci, jak to zostało zaproponowane w jego monografii. Na przykład wykresy typu „gałęzie i liście” („stem-and-leaf”) nie wytrzymują konkurencji ze współczesnymi histogramami, ale wykresy skrzynkowe (box-plot) zrewolucjonizowały prezentację statystyczną. Dodatkowym czynnikiem, który wpłynął na efektywne wykorzystanie metod eksploracji danych, jest rozwój techniki obliczeniowej. Współczesne pakiety statystyczne oferują całą gamę graficznej prezentacji wyników pomiarów, z których duża część ma swoje źródło w prezentacjach Tukeya. Wykorzystanie pięciu statystyk porządkowych⁴ i ich prezentacja graficzna pozyskała dodatkowe określenia⁵ i nowatorską współczesną oprawę graficzną. Idea pozostaje jednak przy koncepcji Tukeya.

J.W. Tukey był genialnym dzieckiem, genialnym studentem i genialnym naukowcem. Przez 40 lat pracował w Bell Telephone Laboratories i jednocześnie przez tych 40 lat był członkiem wspólnoty akademickiej Princeton. Jego myśl widać we wszystkich znaczących osiągnięciach matematycznych, poczynając od Pewnika Wyboru, a kończąc na procedurze FFT i analizie spektralnej⁶. Współuczestniczył w niemal wszystkich znaczących projektach badawczych tamtych czasów – począwszy od badań nad bombą wodorową⁷ i prac nad konstrukcją samolotów U2, poprzez analizę raportu Kinsey’a, udział w opracowaniu spisów ludności, wkład w podstawy rozwoju komputerów i wiele, wiele innych. Jego wyniki badawcze, jak zresztą wielu statystyków, mają charakter pomocniczy i na ogół są prezentowane w postaci wyników grupowych. Pozostawał w przyjaźni z największymi umysłami końca XX wieku, poczynając od Feynmana, Ulama, Halmosa czy Fellera. Trudno przecenić jego wpływ na rozwój statystyki. W pracy [4] D.R. Brillinger podaje 72 sformułowania statystyczne, które uzyskały nowe znaczenie w wyniku wpływu podejścia J.W. Tukeya, między innymi, takie jak „ANOVA”, „jackknife”, „software”, „Lemat Zorna”.

To symboliczne zaprezentowanie sylwetki J.W. Tukeya należy zderzyć z analizą eksploracyjną. Tukey proponuje prostą, żeby nie powiedzieć prymitywną analizę jako fundament wnioskowania. Jest w tym coś, co cechuje wielkich ludzi. Ma świadomość, że dla ogromnej rzeszy badaczy, nie zawsze genialnych i często wykorzystujących statystykę jako narzędzie, należy mówić obrazkami, na których znajduje się prezentacja zagregowanych danych w postaci kropek, kwadracików czy kreseczek. Wiele osób stosowało i stosuje taką wstępną analizę, ale trzeba było wielkiej osobowości Tukeya, żeby świat nauki zaakceptował to podejście jako wartościowe działanie badawcze. Nie stało to się od razu i do dzisiaj mamy pewien opór w akceptacji opracowania wyników pomiarów, w których nie przeprowadzono procedury testowania (w szczególności w publikacjach medycznych). Nie zawsze słusznie.

⁴ Statystyki Tukeya: Min , Q_{25} , Q_{50} , Q_{75} , Max .

⁵ Na wykresach skrzynkowych prezentujemy dodatkowo pomiary „odstające” i „ekstremalne”.

⁶ Fast Fourier Transform.

⁷ Tukey uczęszczał na wykłady Ulama z topologii, a potem już w Princeton blisko współpracowali.

Szereg publikacji, w szczególności raportów dotyczących gospodarki, rozwoju biznesu czy produkcji firmy, nie potrzebuje testowania. Celem tych raportów jest prezentacja obserwacji bez wchodzenia w strukturę losową przedmiotu obserwacji. Prezentacje przeznaczone są dla szerokiego kręgu odbiorców, zatem muszą być dla nich zrozumiałe i czytelne. Stąd mamy niezwykłą wprost popularność metod analiz wskaźników giełdowych opisanych w książce J.J. Murphy'ego [10]. Metody zawarte w podręczniku Murphy'ego nie opierają się na strukturze losowej. Polegają na wyobrażeniu zachowania się wartości finansowych w postaci wykresów wzbogaconych o linie ograniczeń, trendów czy prezentację zmienności. Przypomnijmy, że w 2003 roku Nagrodę Nobla w zakresie ekonomii otrzymali Robert F. Engle i Clive W.J. Granger. Autorzy ci, w swoich analizach statystycznych cen na giełdzie opierają się na wyjątkowo zaawansowanych modelach losowych i ich wyniki stanowią najbardziej ewidentne zastosowanie CDA. Jednak „świat finansowy” ogląda wykresy świecowe lub fale Elliotta, które są w sferze opracowań EDA i te wyniki są podstawą niezliczonej ilości częściowych decyzji finansowych.

Na koniec powróćmy do metodologii badań naukowych. Metodę EDA należy lokować w sferze myśli wywodzących się od Sokratesa⁸.

Jest to typowa postać indukcji rozumianej jako wnioskowanie od „szczegółu” do „ogółu”. Dane empiryczne traktujemy jako przesłanki szczegółowe, a wnioski statystyczne dotyczące analizowanego zjawiska badawczego określamy jako „ogół”. Na przykład pod filozoficznym pojęciem „ogółu” możemy rozumieć sentencję „cena waloru wzrośnie”.

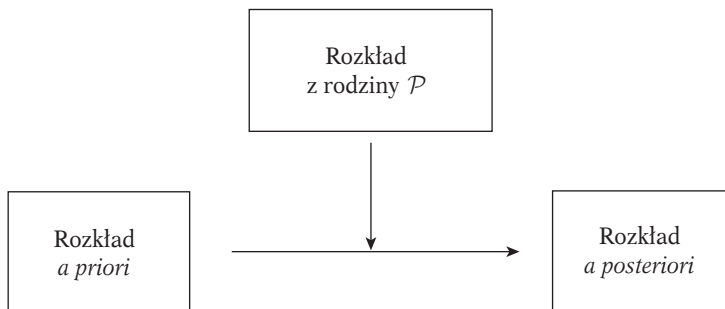
Biorąc pod uwagę, jak Sokrates w Dialogach platońskich dochodził do definicji danego „przedmiotu”, „cechy”, zauważamy zbieżność z analizą eksploracyjną. Stanowisko badawcze określone na przykład w sformułowaniu „cena waloru wzrośnie”, jest konsekwencją wyabstrahowania analogii z szeregu obserwowanych zjawisk (zjawisk dotyczących dotychczasowego zachowania się rynków finansowych). Formalnie rozumując, tak właśnie postępujemy. Dysponujemy zbiorem wyników badawczych i poprzez analizę tych wyników wnioskujemy o konsekwencjach badawczych (patrz [12]).

3. DEDUKCJA – ESTYMACJA BAYESOWSKA

Estymacja bayesowska wymaga znacznej wiedzy matematycznej i intuicji badawczej. Jest najtrudniejsza w stosowaniu, ale chyba najbardziej efektywna. Zobowiązuje do najwyższego poziomu dyscypliny i solidności badawczej.

Samą procedurę postępowania można zapisać w postaci schematu przedstawionego na rysunku 2. Podejście Bayesa do estymacji ma charakter decyzyjny.

⁸ „Sokrates słusznie poszukiwał istoty, poszukiwał bowiem sylogizmów i zasady sylogizmów, bo «to, czym rzecz jest» [istota], jest punktem wyjściowym sylogizmów. [...] Dwa bowiem odkrycia można słusznie przypisać Sokratesowi: wnioskowanie indukcyjne i ogólną definicję, jedno i drugie dotyczy początku wiedzy” (Arystoteles, *Metafizyka*, A 6, 987 b 1 nn).



Rysunek 2. Schemat procedury Bayesa

Źródło: wyniki własne [7].

Podobnie jak w przypadku estymacji punktowej, próba losowa prosta pochodzi z rozkładu z rodziny rozkładów $\mathcal{P}$. W przypadku podejścia Bayesa, zawsze rozkłady z rodziny $\mathcal{P}$ są indeksowane parametrem. Jest to więc sparametryzowana rodzina rozkładów. Akcent badawczy jest przeniesiony z wyboru rozkładu z rodziny $\mathcal{P}$ na wybór parametru. Posługując się prostym przykładem powiemy, że dążymy do oszacowania wartości oczekiwanej z próby losowej prostej, czy znajdujemy rozkład z rodziny rozkładów o wybranej wartości oczekiwanej. W sensie praktycznym są to identyczne informacje, ale w sensie teoretycznym rozpatrujemy różne przestrzenie. W pierwszym przypadku jest to zbiór parametrów, zaś w drugim jest to zbiór rozkładów. Z tym problemem spotykamy się już w przypadku estymacji punktowej, jednak w podejściu bayesowskim rozróżnienie zbioru parametrów i rodziny rozkładów jest daleko bardziej istotne. Można przyjąć, że skupiamy się głównie na parametryzacji rodziny. Oczywiście, nadajemy jej sens empiryczny i staramy się szczegółowo zinterpretować znaczenie i wartość parametru identyfikującego rodzinę.

Kluczowym założeniem w metodzie Bayesa jest nadanie zbiorowi parametrów struktury probabilistycznej. Krótko mówiąc, przyjmujemy, że parametr jest zmienną losową (wektorem losowym). Rozpatrzmy przykład. Niech eksperyment badawczy polega na obserwacji 100 klientów supermarketu, z których każdy zakupi bądź nie zakupi produkt z określonej grupy produktów. Przypisujemy naszemu eksperymentowi rodzinę rozkładów dwumianowych z parametrem $\theta \in (0, 1)$. Parametr ten ma jasno określoną interpretację. Odczytujemy go jako szanse na to, że każdy klient kupuje produkt z częstością θ . W zakresie szkoły średniej w takim przypadku mówimy o prawdopodobieństwie sukcesu. Zawsze jednak mamy fragmentaryczną lub częściową wiedzę na temat parametru. Często proces badawczy zaczyna się od obserwacji wstępnych czy badań pilotażowych, które w konsekwencji wnoszą pewną wiedzę na temat parametru odpowiadającego szansie zakupu. Wiedza taka zostaje potraktowana jako informacja wstępna i skondensowana w postaci rozkładu probabilistycznego, jakiemu podlega cecha losowa. Bardziej formalnie rozkład ten jest rozkładem zmiennej losowej (cechy losowej) Θ nazywanym rozkładem *a priori* (por. [8]).

Zaczynamy więc analizę eksperymentu od dwóch kluczowych założeń. W pierwszym założeniu przyjmujemy określoną rodzinę rozkładów $\mathcal{P}$ indeksowaną parametrem. W drugim założeniu precyzujemy rozkład parametru. Zarówno przyjęcie rodziny rozkładów, jak i rozkładu parametru są subiektywne i wpływają z profesjonalności i merytorycznej wiedzy badacza na temat eksperymentu. W wyborach tych jednak nie można wszystkiego maksymalnie upraszczać. Nie należy również niepotrzebnie popisywać się komplikacją wybranych rozkładów. Podręczniki opisujące podejście bayesowskie podają szereg standardów pozwalających na skuteczne opisanie klasycznych eksperymentów badawczych.

Po wypełnieniu dwóch wyjściowych postulatów, dalsze postępowanie jest automatyczne. Wynikiem estymacji uzyskanym na bazie wyników pomiarów jest rozkład parametru nazwany rozkładem *a posteriori*, który otrzymujemy jako rozkład warunkowy parametru (jako zmiennej losowej Θ) pod warunkiem realizacji próby losowej prostej.

Podkreślmy, że procedura bayesowska polega na tym, że rozkład *a priori* („nasza wiedza”) podlega transformacji w rozkład *a posteriori* („wiedza modyfikowana pomiarami”), w której rolę moderującą odgrywa wyjściowa rodzina rozkładów $\mathcal{P}$ (patrz rysunek 2). Ponieważ wynikiem estymacji Bayesa jest rozkład probabilistyczny, więc we wnioskach badawczych analizujemy charakterystyki tego wynikowego rozkładu. Najczęściej jest to wartość oczekiwana⁹, jednak ponieważ wynikiem jest konkretny jednoznaczny rozkład probabilistyczny, więc prezentowana jest po prostu gęstość tego rozkładu (histogram). Dodatkowo podajemy kwantyle lub charakterystyki rozrzutu.

W sensie teoretycznym procedura jest do zaakceptowania i wydaje się bardzo efektywna. Problemem jest proces obliczeniowy. Jak wiadomo znalezienie rozkładu warunkowego jest na ogół bardzo skomplikowane. Właściwie poza prostymi standardowymi przykładami (które zazwyczaj także wymagają dużej wiedzy z zakresu analizy matematycznej) znalezienie rozkładu *a posteriori*, a czasami tylko jego charakterystyk jest niemal niemożliwe. Z pomocą w tych trudnościach przyspieszyli specjaliści z zakresu klasycznego rachunku prawdopodobieństwa i technik komputerowych. Powstała cała teoria nazwana MCMC (Markov Chain Monte Carlo) pozwalająca na wykorzystanie symulacji komputerowych do otrzymania aproksymacji wynikowego rozkładu *a posteriori* (patrz na przykład [13]).

Na koniec zwróćmy uwagę na naturę filozoficzną koncepcji bayesowskiej. Jest to egzemplifikacja rozumowania Euklidesa, nazwanego dedukcją (patrz na przykład [6]). W podejściu Euklidesa przyjmujemy założenia i na bazie reguł logicznych wysnuwamy konsekwencje w postaci wniosków. W procesie estymacji bayesowskiej założenia obejmują wyróżnienie rodziny rozkładów i rozkładu *a priori*, wnioskiem jest rozkład *a posteriori*.

⁹ W wielu monografiach to właśnie tę wartość oczekiwaną (warunkową) nazywa się estymatorem bayesowskim.

4. PODSUMOWANIE

Jak widać estymacja jest procesem badawczym głęboko osadzonym w filozofii nauki. Spór filozoficzny na temat metodologii nauki trwa setki lat i będzie trwał nadal. Tak samo w estymacji nie powiemy, która metoda jest lepsza, która gorsza. W procedurze doboru sposobu estymacji decyduje sumienność naukowa badacza, profesjonalność w zakresie meritum zagadnienia i zakres wynikowych wnioskowań. Niektórzy badacze kierowani specyfiką przedmiotu badań czy osobistym nastawieniem będą preferować jedno czy drugie podejście, ale nie powinno to przeważać w wyborze metodologii estymacji.

LITERATURA

- [1] Ajdukiewicz K., [1955], *Klasyfikacja rozumowań*, *Studia Logica*, II:278-300.
- [2] Barra J.R., [1982], *Matematyczne podstawy statystyki*, Państwowe Wydawnictwo Naukowe, Warszawa.
- [3] Bickel P.J., Doksum K.A., [2001], *Mathematical Statistics. Basic Ideas and Selected Topics*, Vol. I, Prentice Hall, Upper Saddle River, New Jersey, 07458, wydanie II.
- [4] Brillinger D.R., Tukey J.W., [2002], *His life and professional contributions*, *The Annals of Statistics*, 30(6):1535-1575.
- [5] Gelman A., Carlin J.B., Stern H.S., Rubin D.B., [2000], *Bayesian Data Analysis*, CHAPMAN & HALL/CRC, wydanie II.
- [6] Herbut J., (redaktor), [1997], *Leksykon filozofii klasycznej*, Katolicki Uniwersytet Lubelski, hasło DEDUKCJA – autor hasła Andrzej Bronk, s. 101-102.
- [7] Krzykowski G., [2000], *Analiza wyników badań marketingowych. Metody symulacyjne*, Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk.
- [8] Krzykowski G., [2010], *Pomiar i statystyczna analiza cech binarnych*, Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk.
- [9] Lehmann E.L., [1991], *Teoria estymacji punktowej*, Państwowe Wydawnictwo Naukowe, Warszawa.
- [10] Murphy J.J., [1999], *Analiza techniczna rynków finansowych*, WIG-PRESS.
- [11] Petras I., Podlubny I., [2010], *Least Squares or Least Circles? A comparsion of classical regression and orthogonal regression*, *Chance*, 23(2):38-42.
- [12] Reale G., *Historia filozofii starożytnej*, Katolicki Uniwersytet Lubelski, 2000-2002.
- [13] Robert C.P., Casella G., [2004], *Monte Carlo Statistical Methods*, Springer-Verlag, wydanie II.
- [14] Stigler S.M., [1981], *Gauss and invention of last squares*, *The Annals of Statistics*, 9(3):465-474.
- [15] Tukey J.W., [1977], *Exploratory Data Analysis*, Addison-Wesley.
- [16] Woleński J., [1985], *Filozoficzna szkoła lwowsko-warszawska*, Państwowe Wydawnictwo Naukowe, Warszawa.

POLSCY STATYSTYCY I MATEMATYCY

PROFESOR MAREK FISZ (1910-1963). W SETNĄ ROCZNICĘ URODZIN



Marek Fisz urodził się 15 stycznia 1910 roku w Szydłowcu w rodzinie żydowskiej. Rodzice, Sura, z domu Rozenbaum, i Kojfman Fisz mieszkali w Ostrowcu Świętokrzyskim. Ojciec był prywatnym urzędnikiem (pracował między innymi jako pisarz w miejscowym młynie). Pierwotne imię Marka Fisza brzmiało Mojżesz. Zmiana imienia na Marek nastąpiła 19 lipca 1950 roku decyzją Prezydium Dzielnicowej Rady Narodowej Warszawa-Zachód. Dzieciństwo i młodość Marka Fisza owiane są tajemnicą. Wiadomo tylko, że w maju 1934 roku uzyskał świadectwo dojrzałości jako eksternista. W życiorysie dołączonym do wniosku o wszczęcie Mu przewodu doktorskiego napisał: „Do przerabiania kursu szkoły średniej przystąpiłem w 22-im roku życia”. Jesienią 1934 roku rozpoczął studia matematyczne na Wydziale Matematyczno-Przyrodniczym Uniwersytetu Warszawskiego. W okresie studiów, w latach 1936-1939, pracował jako nauczyciel matematyki w szkole dla dorosłych prowadzonej przez Stowarzyszenie Pomocy Wzajemnej Lokatorów „Szklane Domy” przy Warszawskiej Spółdzielni Mieszkaniowej na Żoliborzu. Pracę magisterską pt. „Konformiczne przekształcenia przestrzeni jedno- i dwuspójnych” napisał pod kierunkiem profesora Stanisława Saksa. Po zdaniu w dniu 28 czerwca 1939 roku egzaminu magisterskiego z wynikiem dobrym, uzyskał tytuł magistra filozofii w zakresie matematyki.

Przez cały okres wojny przebywał na terenie Związku Radzieckiego. 29 stycznia 1944 roku ożenił się z Olgą Gukow. Mieli jednego syna Aleksandra, urodzonego 24 września 1946 roku.

W roku 1946 Marek Fisz wrócił do Polski i przez rok pracował jako wychowawca w Domu Dziecka w Otwocku. Od jesieni 1947 roku do końca września 1951 roku pracował w Głównym Urzędzie Statystycznym w Warszawie w charakterze pracownika naukowego. Warto w tym miejscu zauważyć, że z pracą naukową matematyka zetknął się po raz pierwszy mając ponad 37 lat. W szczególności, kierował tam zaprojektowaniem i opracowaniem wyników pierwszego po II wojnie światowej Narodowego Spisu Powszechnego, przeprowadzonego według stanu o północy z 2 na 3 grudnia 1950 roku.

20 maja 1950 roku, Marek Fisz zwrócił się do Rady Wydziału Matematyki, Fizyki i Chemii Uniwersytetu i Politechniki we Wrocławiu z następującą prośbą: „Uprzejmie proszę o przyjęcie mojej pracy „Kontrola jakości produkcji masowej na cechę ciągłą” jako pracy doktorskiej i, w razie uznania jej za wystarczającą o dopuszczenie mnie do egzaminu ścisłego celem uzyskania stopnia doktora filozofii w zakresie matematyki (przedmiot uboczny: mechanika teoretyczna). Ścisły egzamin doktorski Marek Fisz zdał 19 grudnia 1950 roku przed Komisją w składzie: prof. dr Jan Nikliborc – przewodniczący oraz prof. dr Hugo Steinhaus, prof. dr Edward Marczewski, prof. dr Władysław Ślebodziński – członkowie. Promotorem rozprawy doktorskiej pt. „Kontrola jakości produkcji masowej na cechę ciągłą” opublikowanej w czasopiśmie *Studia i Prace Statystyczne*, z.2(1950), 123-160, był profesor Hugo Steinhaus. Jego promocja doktorska odbyła się 23 czerwca 1951 roku w sali Senatu Uniwersytetu Wrocławskiego.

L. 959/51 (220) Wydział: Mat-Fiz-Chem L. p. Wyd. 18
wpłynęło dnia 25.V.1950

P R O T O K Ó Ł P R O M O C J I D O K T O R S K I E J

Imię i nazwisko Doktoranda: Marek Fisz
Data urodzenia i miejscowość: 15.I.1910 Szydówko Waj. Kieleckie
Wyznanie: _____ Narodowość: polska

O d b y ł o s t u d i a :

I rok	Wydział Mat-Fiz-Chem	Data
II	Uniw. Warszawski	1935
III	"	1936
IV	"	1937
V	"	"

Praktyka odbyta w: /
Tytuł pracy magisterskiej lub dyplomowej: Przebiegi i konformacje przestrzeni jedno- i dwuwymiarowych
Data egzaminu magist. lub dyplom.: 28.VI.1939 Wynik: dobry
Uczelnia: Uniwersytet Warszawski
Tytuł pracy doktorskiej: "Kontrola jakości produkcji masowej na cechę ciągłą"
drukowanej: Kontrola jakości produkcji masowej na cechę ciągłą
Promotor: Prof. Dr. Hugo Steinhaus
Dziękuję: „Władysław Ślebodziński”
Rektor: „Stanisław Kalonizki”
Data promocji: 23.VI.1951 godz. 13-ta
Miejsce promocji: Uniwersytet Wrocławski
Nr. Dyplomu Doktorskiego: _____
Przyrzeczenie doktorskie złożyłem: _____
Doktorand: _____
Promotor: _____
Rektor U. i P.: _____
Wrocław, dnia _____ mies. _____ 194 r.

Od czerwca 1950 roku rozpoczął współpracę z Działem Statystyki Matematycznej w Państwowym Instytucie Matematycznym w Warszawie, a od października 1951 roku podjął pracę w tym Instytucie. Jednocześnie, od roku akademickiego 1951/52 rozpoczął współpracę z Uniwersytetem Warszawskim, prowadząc tam wykłady i seminaria zlecone na Wydziale Matematyki, Fizyki i Chemii. Z dniem 1 kwietnia 1954 roku minister Szkolnictwa Wyższego powołał Go na kierownika Katedry Statystyki Matematycznej na Wydziale Matematyki, Fizyki i Chemii Uniwersytetu Warszawskiego. Od tego momentu stał się stałym pracownikiem tego Wydziału. Centralna Komisja

Kwalifikacyjna dla Pracowników Nauki uchwałą z dnia 30 czerwca 1954 roku przyznała Drowi Markowi Fiszowi tytuł naukowy docenta. Ta sama Komisja uchwałą z dnia 28 marca 1957 roku przyznała Mu tytuł naukowy profesora nadzwyczajnego. Sekcja Studiów Uniwersyteckich Rady Głównej Szkolnictwa Wyższego na posiedzeniu w dniu 20 marca 1957 r. powołała Go na członka Zespołu Rzeczoznawców Matematyki. W dniu 15 marca 1958 roku został powołany na kierownika Działu Statystyki Matematycznej w Instytucie Matematycznym PAN w Warszawie.

Brał udział w I Kongresie Nauki Polskiej, który trwał od 29 czerwca do 2 lipca 1951 roku w Warszawie. Był to ogólnopolski zjazd około 1800 naukowców. Powołano na nim Polską Akademię Nauk. Od roku 1949 był członkiem Polskiego Towarzystwa Matematycznego, a w latach 1953-1955 był skarbnikiem Zarządu Głównego tegoż Towarzystwa. Uczestniczył w organizacyjnym zjeździe Sekcji Biometrycznej przy Polskim Towarzystwie Przyrodników im. Mikołaja Kopernika, który odbył się w dniach 19-21 lutego 1959 roku w Instytucie Matematycznym PAN we Wrocławiu z udziałem J. Czekanowskiego, B. Knastera, J. Łukaszevicza, W. Oktaby, M. Olekiewicza, J. Perkala, H. Steinhausa, S. Zubrzyckiego i innych. Na początku 1961 roku Sekcja ta przekształciła się w Polskie Towarzystwo Biometryczne, którego pierwszym prezesem został Julian Perkal.

Profesor Marek Fisz wygłosił referaty między innymi na:

- Ogólnopolskim Zjeździe Matematyków w Warszawie w roku 1953
- Konferencji probabilistycznej w Berlinie w roku 1954
- Konferencji poświęconej procesom stochastycznym we Wrocławiu w roku

1954

- Konferencji probabilistycznej w Leningradzie w roku 1955
- Czwartym Sympozjum w Berkeley w roku 1960
- Dorocznym Spotkaniu Instytutu Statystyki Matematycznej w Seattle w roku

1961.

W roku 1955, Profesor Marek Fisz spędził trzy miesiące na Uniwersytecie Moskiewskim, a wiosną 1957 roku wykładał na Uniwersytecie Pekińskim.

Czynił starania o wyjazd na stałe do Izraela, ale nie uzyskał zgody władz polskich.

W roku 1960 wyjechał na roczny staż naukowy do Stanów Zjednoczonych, z którego nie powrócił do Polski. Pracował w University of Washington (Seattle), w Stanford University (Stanford), w Columbia University oraz w New York University (New York).

Zmarł 4 listopada 1963 roku w Metropolitan Hospital of New York City, U.S.A., mając niewiele ponad 53 lata. Jest rzeczą zmienną, że żadne polskie czasopismo naukowe nie odnotowało śmierci Profesora chociażby jednym zdaniem.

W krótkim, bo tylko szesnastoletnim okresie Jego kariery naukowej uzyskał wiele znaczących wyników. Pierwsze prace naukowe Profesora Marka Fisz, włącznie z rozprawą doktorską [4], dotyczyły metod pobierania prób oraz kontroli jakości produkcji ([1], [2], [3], [5], [6], [14], [18]). Kolejną grupę stanowią prace dotyczące rozkładów granicznych funkcji zmiennych losowych dyskretnych, ze szczególnym uwzględnieniem rozkładu Poissona i rozkładu wielomianowego ([7], [8], [9], [11], [17], [19], [20], [20], [21]). Nowoczesność badań Profesora Marka Fisz uwidoczniła się w serii

Jego artykułów z lat 1957-1958 ([25], [26], [27], [32]) dotyczących rozkładów granicznych statystyk nieparametrycznych typu Kołmogorowa-Smirnowa. W artykułach tych wykorzystywał wynik Prochorowa z roku 1956 dotyczący słabej zbieżności miar do wielowymiarowych procesów empirycznych. Wyniki te nawet dzisiaj nadają się do pełnego wykorzystania w statystyce matematycznej. Artykuły te zostały przygotowane w czasie Jego wizyt w Moskwie i Pekinie. Profesor Marek Fisz znaczącą uwagę poświęcił również badaniom własności funkcji próby procesów stochastycznych. W pracy [22] podał warunki wyrażone w terminach prawdopodobieństw bezwzględnych, przy których prawie wszystkie funkcje próby procesu są funkcjami skokowymi ze skończoną oczekiwaną liczbą punktów nieciągłości. Wyniki te znacząco uogólnił w artykule [39], który zaprezentował na Czwartym Sympozjum w Berkeley w roku 1960. 16 czerwca 1961 roku w Seattle w Stanie Washington, Profesor Marek Fisz wygłosił specjalny referat zaproszony [40] na Dorocznym Spotkaniu Instytutu Statystyki Matematycznej. Przedmiotem Jego wystąpienia były rozkłady nieskończone podzielne, którym pozostał wierny aż do śmierci (np. [42], [43], [44], [45]). Pyke (1967) napisał: „This excellent expository address [40], together with its extensive bibliography of 104 references will continue to be very useful for many years to come”.

Najbardziej znanym wkładem Profesora Marka Fisza do teorii prawdopodobieństwa i statystyki matematycznej jest Jego podręcznik *Rachunek prawdopodobieństwa i statystyka matematyczna*, którego pierwsze wydanie ukazało się w roku 1954 i liczyło 374 strony tekstu. Profesor Marek Fisz pracę nad tym podręcznikiem rozpoczął w roku 1950, zanim uzyskał doktorat. Podręcznik ten okazał się bardzo popularny i wszystkie egzemplarze zostały sprzedane w ciągu paru miesięcy. W roku 1958 nakładem Państwowego Wydawnictwa Naukowego ukazało się wydanie drugie tego podręcznika, poprawione i rozszerzone jako Tom 18 Biblioteki Matematycznej. Wydanie to liczyło już 530 stron. Podręcznik Profesora Marka Fisza na wiele lat wyznaczył standardy nauczania rachunku prawdopodobieństwa i statystyki matematycznej. Wydanie trzecie tego podręcznika w języku angielskim ukazało się w wydawnictwie Wiley w roku 1963. Tłumaczem z języka polskiego był Profesor Robert Bartoszyński. Nawiasem mówiąc, promotorem rozprawy doktorskiej Roberta Bartoszyńskiego pt. „O słabej zbieżności miar” obronionej 23 stycznia 1960 roku w Instytucie Matematycznym PAN w Warszawie był Profesor Marek Fisz. Z kolei Profesor Robert Bartoszyński był jednym z opiniodawców habilitacji i profesury autora tego artykułu. Wydanie trzecie podręcznika Marka Fisza w języku polskim, nakładem Państwowego Wydawnictwa Naukowego, ukazało się dopiero w roku 1967, już po śmierci Autora. Opóźnienie to było wynikiem zawirowań życiowych Marka Fisza. Wydanie to liczyło 694 strony. Pyke (1967) o podręczniku tym napisał tak: „This text was written for either senior undergraduate or first-year graduate students. It covers both probability theory and statistical inference, and includes introductions to stochastic processes and sequential analysis in addition to the more usual topics. There is a large collection of 317 problems and complements in the book, which, together with its considerable historical and bibliographical information, make it an excellent reference book as well as a textbook”. Oprócz czterech wydań tego podręcznika w języku polskim ukazały się trzy wydania w języku angielskim (jedno w Indiach i dwa w USA) oraz jedenaście wydań w języku niemieckim.

Profesor Marek Fisz w ciągu stosunkowo krótkiej kariery naukowej wniósł bardzo wiele do rozwoju i nauczania zarówno rachunku prawdopodobieństwa, jak i statystyki matematycznej.

Mirosław Krzyśko
Wydział Matematyki i Informatyki Uniwersytetu im. Adama Mickiewicza
i Instytut Zarządzania Państwowej Wyższej Szkoły Zawodowej
im. Prezydenta Stanisława Wojciechowskiego w Kaliszu

BIBLIOGRAFIA TWÓRCZOŚCI PROFESORA MARKA FISZA

ARTYKUŁY

1949

1. O reprezentacji literowej populacji ludnościowej (Letter sampling in population investigation), *Przegląd Statystyczny* 3 (1949), 81-112 (Polish).

1950

2. Projekt badania reprezentacyjnego spisu powszechnego (Project of a sample survey of the census material), *Studia i Prace Statystyczne* 1 (1950), 50-59 (Polish).
3. Program prac statystycznych na 1950 r. Komisja do spraw statystyki matematycznej (Statistical work program guides for 1950. The Committee on mathematical statistics), *Studia i Prace Statystyczne* 1 (1950), 96-97 (Polish).
4. Kontrola jakości produkcji masowej na cechę ciągłą (Quality control of mass production by variables), *Studia i Prace Statystyczne* 2 (1950), 123-160 (Polish).
5. Z prac Komisji do spraw statystyki matematycznej przy GUS. Konsultacje z prof. J. Neymanem i wnioski z nich wypływające (From the work of the Committee on mathematical statistics at the Central Statistical Office. Consultations with prof. J. Neyman and conclusions following from them), *Studia i Prace Statystyczne* 3/4 (1950), 14-27 (Polish).

1951

6. Efektywność dwóch metod pobierania prób (The efficiency of two sampling methods), *Studia i Prace Statystyczne* 1 (1951), 91-93 (Polish).

1953

7. Rozkład graniczny różnicy dwóch zmiennych losowych o rozkładzie Poissona (The limiting distribution of the difference of two Poisson random variables), *Zastosowania Matematyki* 1 (1953), 41-45 (Polish).
8. The limiting distributions of sums of arbitrary independent and equally distributed r -point ($r \geq 2$) random variables, *Bull.Acad.Pol.Sci.,Cl III*, 1(1953), 235-238 (English).

1954

9. Die Wahrscheinlichkeitsrechnung als Hilfsmittel zu Untersuchungen in Naturwissenschaften und Production (with Steinhaus H., Czechowski T., Lange O., Oderfeld J., Sadowski W.), *Hauptreferate 8.Poln.Math.-Kongr.*, 6-12.Sept.1953 Warschau, 69-94 (1954) (German).
10. Dokładność pewnego wzoru asymptotycznego (Accuracy of an asymptotical formula), *Zastosow. Mat.*2(1954), 62-66 (Polish).
11. The limiting distributions of sums of arbitrary independent and equally distributed r -point ($r \geq 2$) random variables, *Stud.Math.*14(1954), 111-123 (English).

12. The limiting distributions of the multinomial distribution, *Stud.Math.*14(1954), 272-275 (English).
13. A generalization of a theorem of Kintchin, *Stud.Math.*14(1954), 310-313 (English).

1955

14. O wyznaczaniu współczynnika bezpieczeństwa (On determining the safety factor) (with Czechowski, T., Sadowski, W., Zasepa, R.), *Zastosow.Mat.*2(1955), 190-198 (Polish).
15. The limiting distribution of a function of two independent random variables and its statistical application, *Colloq.Math.*3(1955), 138-146 (English).
16. Refinement of a probability limit theorem and its application to Bessel functions, *Acta Math. Acad. Sci. Hung.*6(1955), 199-202 (English).
17. The analytical characterization of the composed non-homogeneous Poisson process (with Urbanik, K.), *Bull. Acad. Pol. Sci., Cl. III* 3 (1955), 149-150 (English).
18. Konstrukcja populacji sztucznych i ich zastosowanie (The construction of artificial populations and their application), *Pr.Mat.*1(1955), 174-182 (Polish).

1956

19. A limit theorem for a modified Bernoulli scheme, *Stud. Math.*15(1956), 80-83 (English).
20. Die Grenzverteilungen der Multinomialverteilung, *Ber. Tagung Wahrscheinlichkeitsrechnung Math. Statist., Berlin* 1954, 51-53(1956) (German).
21. The analytical characterization of the composed non-homogeneous Poisson process (with Urbanik, K.), *Stud. Math.*15(1956), 328-336 (English).
22. Realizations of some stochastic processes, *Stud. Math.*15(1956), 359-364 (English).
23. The realizations of some purely discontinuous stochastic processes, *Bull. Acad. Pol. Sci., Cl. III* 4 (1956), 63-65 (English).
24. Uwagi o rachunku prawdopodobieństwa, w: *Statystyka jako metoda poznawcza* (red. T. Czechowski i inni) (Remarks on probability theory, In: *Statistics as a cognitive method* (red. T. Czechowski et al.), *Zeszyty problemowe „Kosmosu”, z. 2*, PWN, Warszawa 1956, 7-20 (Polish).

1957

25. Asymptotically independent linear functions of empirical distribution functions (with Chang, L., C.), *Science Record, n. Ser.* 1(1957), 335-340 (English).
26. Exact distributions of the maximal values of some functions of empirical distribution functions (with Chang, L., C.), *Science Record, n. Ser.*1(1957), 341-346(English).
27. A limit theorem for empirical distribution functions, *Bull. Acad. Pol. Sci., Cl. III* 5 (1957), 695-698 (English).
28. Kharakteristicheskiye svoystvo normal'nogo raspriyedel'eniya v gil'byertovom prostranstve (A characterization of normal distributions in Hilbert spaces) (with Yu. V. Prohorov), *Theor. Verojatn. Primen.* 2(1957), 475-477 (Russian).

1958

29. A central limit theorem for some stochastic processes, *Bull. Acad. Pol. Sci., Ser. Sci. Math. Astron. Phys.* 6(1958), 437-443 (English).
30. A limit theorem for non-decreasing random functions, *Bull. Acad. Pol. Sci., Ser. Sci. Math. Astron. Phys.* 6 (1958), 485-487 (English).
31. Characterization of some probability distributions, *Skand. Aktuarietidskr.*41 (1958), 65-67 (English).
32. A limit theorem for empirical distribution functions, *Stud. Math.* 17(1958), 71-77 (English).

1959

33. A central limit theorem for stochastic processes with independent increments, *Stud. Math.* 18(1959), 223-227 (English).
34. On necessary and sufficient conditions for the validity of the strong law of large numbers expressed in terms of moments, *Bull. Acad. Pol. Sci., Ser. Sci. Math. Astron. Phys.* 7(1959) (English).

1960

35. Remarks on the sample functions of some stochastic processes, Bull. Acad. Pol. Sci., Ser. Sci. Math. Astron. Phys. 8(1960), 355-358 (English).
36. On a result by M. Rosenblatt concerning the von Mises-Smirnov test, Ann. Math. Stat. 31(1960), 427-429 (English).
37. Some non-parametric tests for the k-sample problem, Colloq. Math. 7(1960), 289-296 (English).
38. A necessary and sufficient condition of the validity of the weak law of large numbers (with A. Ehrenfeucht), Bull. Acad. Pol. Sci., Ser. Sci. Math. Astron. Phys. 8 (1960) (English).

1961

39. Characterization of sample functions of stochastic processes by some absolute probabilities, Proc. 4th Berkeley Symp. Math. Stat. Probab. 2(1961), 143- 151 (English).

1962

40. Infinitely divisible distributions: Recent results and applications, Ann. Math. Stat. 33(1962), 68-84 (English).
41. On an inequality due to Gatti and Birnbaum, Metron 22(1962), 110-112 (English).
42. On the continuity of the L-distribution functions, Z. Wahrscheinlichkeitstheor. verw. Geb. 1(1)(1962), 25-27 (English).

1963

43. A condition for absolute continuity of infinitely divisible distribution functions(with V. S. Varadarajan), Z. Wahrscheinlichkeitstheorie. verw. Geb. 1(4) (1963), 335-339 (English).
44. On the orthogonality of measures induced by L-processes, Trans. Am. Math. Soc. 106(1963), 185-192 (English).
45. On the absolute continuity of measures induced by stationary L-processes, Unpublished Tech. Rpt. from Columbia University, 15p.(English).
46. On Kolmogorov's continuity condition for sample functions, Unpublished Tech. Rpt. from Columbia University, 6p.(English).

KSIAŻKI

(a) w języku polskim

- (1) Rachunek prawdopodobieństwa i statystyka matematyczna (Probability theory and mathematical statistics), red. Józef Łukasiewicz, Państwowe Wydawnictwo Naukowe, Warszawa 1954, 374 p.
- (2) Tablice statystyczne (opracowali: Tadeusz Czechowski, Marek Fisz, Tadeusz Iwiński, Oskar Lange, Wiesław Sadowski, Ryszard Zasepa, redagował: Wiesław Sadowski) (Statistical tables, prepared by: Tadeusz Czechowski, Marek Fisz, Tadeusz Iwiński, Oskar Lange, Wiesław Sadowski, Ryszard Zasepa, edited by Wiesław Sadowski), Państwowe Wydawnictwo Naukowe, Warszawa 1957, 158 p.
- (3) Rachunek prawdopodobieństwa i statystyka matematyczna, Wydanie drugie poprawione i rozszerzone (Probability theory and mathematical statistics, Second edition revised and expanded), Biblioteka Matematyczna, Tom 18, Państwowe Wydawnictwo Naukowe, Warszawa 1958, 530 p.
- (4) O rachunku prawdopodobieństwa i jego zastosowaniach (On probability theory and its applications), Polskie Towarzystwo Chemiczne, Warszawa 1960.
- (5) Rachunek prawdopodobieństwa i statystyka matematyczna, Wydanie trzecie poprawione i rozszerzone (Probability theory and mathematical statistics, Third edition revised and expanded), Biblioteka Matematyczna, Tom 18, Państwowe Wydawnictwo Naukowe, Warszawa 1967, 694 p.
- (6) Rachunek prawdopodobieństwa i statystyka matematyczna, Wydanie czwarte (Probability theory and mathematical statistics, Forth edition), Biblioteka Matematyczna, Tom 18, Państwowe Wydawnictwo Naukowe, Warszawa 1969, 694 p.

(b) w języku angielskim

- (1) Probability and mathematical statistics (Translated from the Polish 2ed ed. by P. K. Kulshrestha), S.Chand, Delhi 1961, 502p.
- (2) Probability theory and mathematical statistics (Translated from the Polish 3rd ed. by R. Bartoszyński), Wiley, New York 1963, 677 p.
- (3) Probability theory and mathematical statistics (Translated from the Polish 3rd ed. by R. Bartoszyński, repr. with corr.), Robert E. Krieger Publishing Company, Huntington (N. Y.) 1980, 677 p.

(c) W języku niemieckim

- (1) Wahrscheinlichkeitsrechnung und mathematische Statistik (Autorisierte Übersetzung aus dem Polnischen nach der 2., verbesserten und erweiterten Aufl.:Josef Wloka, Redaktion: Hannelore und Rolf Sulanke), Hochschulbücher für Mathematik, Bd. 40, VEB Deutscher Verlag der Wissenschaften, Berlin 1958, 528 S.
- (2) 2.,Aufl.,Berlin 1962, 528 S.
- (3) 3.durchgesehene Aufl., Berlin 1965, 551 S.
- (4) 4.,durchgesehene Aufl., Berlin 1966, 551 S.
- (5) 5.,erw. Aufl., Berlin 1970, 777 S.
- (6) 6. Aufl., Berlin 1971, 777 S.
- (7) 7. Aufl., Berlin 1973, 777 S.
- (8) 8. Aufl., Berlin 1976, 777 S.
- (9) 9. Aufl., Berlin 1978, 777 S.
- (10) 10. Aufl., Berlin 1980, 777 S.
- (11) 11. Aufl., Berlin 1989, 777 S.

ŹRÓDŁA

- Domański Cz., Fisz Marek, 1910-1963, W: Słownik biograficzny statystyków polskich, Główny Urząd Statystyczny i Polskie Towarzystwo Statystyczne, Warszawa 1998, 92-93.
- Oktaba W., Fisz Marek, W: Oktaba W., Probabiliści, statystycy, matematycy, ekonometrycy i biometrycy od starożytności do 2000 r., Lubelskie Towarzystwo Naukowe, Lublin 2002, 98-99.
- Pyke R., Obiturity of Professor Marek Fisz, Z. Wahrscheinlichkeitstheorie verw. Geb. 8 (1967), 153-156.

MATERIAŁY ARCHIWALNE:

1. Instytutu Matematycznego PAN w Warszawie,
2. Uniwersytetu Wrocławskiego,
3. Urzędu Stanu Cywilnego w Szydłowcu,
4. Wydziału Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego.

Zbiory biblioteczne Wydziału Matematyki i Informatyki Uniwersytetu im. Adama Mickiewicza w Poznaniu.

RECENZJE

ANTONI SMOLUK

RECENZJA KSIĄŻKI:
ALEKSANDER JAKIMOWICZ,
ŹRÓDŁA NIESTABILNOŚCI STRUKTUR RYNKOWYCH,
WYDAWNICTWO NAUKOWE PWN SA, WARSZAWA 2010. CENA 49,90 PLN

Książka – oprawiona w zielone miękkie okładki z białymi napisami – ukazała się w serii wydawniczej *Współczesna ekonomia*. Jest – być może – dziełem życia Autora. Napisana językiem pięknym i świeżym, rzadko teraz spotykanym w pracach naukowych; jest wielkim sukcesem Autora i dużym osiągnięciem wydawnictwa.

Na początek dane statystyczne. Praca liczy bez bibliografii 287 stron, do tego dochodzi jeszcze 26 stron bibliografii, dziewięciostronicowy indeks i dwustronicowe streszczenie po angielsku. Łącznie ze spisem rzeczy, stronami reklamowymi wydawnictwa, naturalnym otwarciem i zamknięciem mamy 336-stronicowy utwór. Ciało główne składa się ze wstępu, 13 stron mierzącego, 7 rozdziałów mających łącznie 233 strony, zakończenia na 7 stron i aneksu liczącego 34 strony. Żywa paginacja ułatwia lekturę, jednocześnie zdobi i ożywia przedmiot.

Rozdział I *Teoremat pajęczyny* podzielony na 6 paragrafów, ma 60 stron oraz 22 rysunki. Rozdział II *Teoria wyboru konsumenta; model Rickera* zajmuje 19 stron, ma cztery paragrafy i jest okraszony 8 rysunkami. Rozdział III *Teoria monopolu*, to 7 paragrafów na 43 stronach oraz 17 rysunków. Rozdział IV *Stan badań nad oligopolem* jest bez rysunków i paragrafów, a liczy tylko 15 stron druku. Rozdział następny V *Dynamika duopolu* ma 42 strony druku rozbitego na 5 paragrafów i sztafaż z 18 rysunków. Kolejny rozdział VI *Oligopol z trzema przedsiębiorstwami* obejmuje 38 stron zilustrowanych 10 rysunkami. Wreszcie rozdział ostatni VII *Identyfikacja nieliniowości i chaosu w ekonomicznych szeregach czasowych* liczy stron 16, nie ma paragrafów ani rysunków, ale za to ma cenną tablicę będącą przeglądem literatury na tytułowy temat. Po zakończeniu i aneksie podzielonym na 10 punktów następuje spis literatury – około 510 pozycji, dalej idzie bardzo dobry indeks terminów ułatwiający lekturę, a całość wieńczy angielskie streszczenie.

Część rysunków jest kolorowych; są to często prawdziwe dzieła sztuki abstrakcyjnej. Dotyczy to głównie rysunków z rozdziału III, a szczególnie tych trzech ostatnich z numerami 15, 16 i 17. Część rysunków zaopatrzone w tablicę towarzyszącą, będącą swego rodzaju legendą obrazu. Tablice te kalibrują kolory; rysunek ma bowiem charakter barwnego wykresu rozkładu prawdopodobieństwa – kolory mówią o częstościach pojawienia się trajektorii w wyszczególnionych regionach obszaru zmienności parametrów.

Źródłem tytułowych perturbacji struktur rynkowych są zachowania agentów: producentów co drogo chcą sprzedać i konsumentów, którzy tanio pragną nabyć wybrane wyroby. Ta sprzeczność w warunkach konkurencji niedoskonałej rodzi chaos. Zawodzi niewidzialna ręka rynku. Chaos ekonomiczny nie jest gorszy od chaosu fizycznego, czy kosmicznego. Jego źródłem nie są modele – obojętnie czy liniowe, czy nieliniowe, lecz natura. Rynek jest złożonym systemem, nieprzewidywalnym układem cybernetycznym z powodu irracjonalnych decyzji agentów. Człowiek wprowadza nieokreśloność, losowość, do tego systemu. Ekonomia, bez mglistych preferencji konsumentów, byłaby szwajcarskim zegarkiem. Pełna zgodność równowagi produk-

cyjnej z równowagą rynkową, to czysty komunizm. Preferencje tu są określone, stałe i nie ulegają zmianie. Wyższa instancja wie co ludziom poszczególnym się należy i potrzeba, a wszyscy naturalnie to akceptują. Jak w systemach utopijnych. Jest tu optimum organizacyjne – mrowisko albo ul. Oczywiście świat jest nieliniowy i złożony. Jednak prawa nauki dotyczą obiektów idealnych i obowiązują w określonym czasie. Modele stosowane w praktyce są różne, tak liniowe jak i nieliniowe. Prawo wzrostu naturalnego, słynne równanie Malthusa,

$$y' = ry$$

mówi o przyroście, który jest proporcjonalny do osiągniętego stanu systemu. Jest to równanie różniczkowe liniowe pierwszego rzędu. Jego rozwiązaniem jest funkcja wykładnicza

$$y = a \exp(rt),$$

czyli funkcja nieliniowa. Prawo nauki jest liniowe, trajektorie – nieliniowe. I tak jest często w nauce, a szczególnie w ekonomii; swoistość ta jest cechą charakterystyczną ogólnej teorii zwyczajnych równań różniczkowych liniowych. Jeśli dynamikę systemu złożonego – a takim jest przecież ekonomia społeczna – opisujemy równaniami różniczkowymi i to nawet wyższych rzędów, to zawsze będziemy w sytuacji paradoksalnej: zasady liniowe generują dynamikę nieliniową. Równanie liniowe zwyczajne ma trzy gatunki rozwiązań: wielomiany, funkcje wykładnicze oraz funkcje okresowe – sinusoidalne, a ponadto liniowe kombinacje iloczynów tych trzech typów zależności funkcyjnych. Nauka w ogólności, fizyka w szczególności, a także inne dziedziny oparte i wzorujące się na fizyce, korzysta z równań różniczkowych – zwyczajnych i cząstkowych – tylko rzędu drugiego. Równania różniczkowe drugiego rzędu są fundamentalnym językiem i narzędziem poznania natury. Pierwsza pochodna jest przyrostem, druga inwestycjami, a czym jest trzecia tego już nie wiemy. Przyrostem inwestycji? Jeśli tak, to wracamy znów cyklicznie do pierwszej pochodnej. Z tego co napisałem widać, że pojęcia liniowości i nieliniowości wzajemnie się przeplatają i nie można *en gros* ich traktować oddzielnie. Jednocześnie podkreślić trzeba, że każda funkcja gładka lokalnie jest prawie liniowa. Na tym polega istota rachunku różniczkowego; obiekty nieliniowe zastępuje się lokalnie funkcjami liniowymi optymalnie przylegającymi, czyli stycznymi. Takimi argumentami usprawiedliwia się bogatą i popularną teorię modeli liniowych w ekonometrii. Cena jest funkcjonalem liniowym, jednak przy większym zakupie otrzymujemy rabat, a to znaczy, że w detalu mamy liniowość, a w hurcie subliniowość. To kolejny argument za nierozzerwalnym związkiem obiektów liniowych i nieliniowych. Relacje liniowe są w praktyce rzadkością. Często je się przyjmuje dla uproszczenia. Rozwój naturalny – jak powiedziano wyżej – jest wykładniczy, a więc nieliniowy. To stwierdzenie zdecydowanie wzmacnia argumenty Autora, a ponadto model pajęczynowy – podstawowa idea obrazująca dążenie do równowagi – jest ruchem wirowym, czyli spiralą, a więc tworem nieliniowym. Ten ruch wirowy wokół punktu równowagi jest podstawą ekonomii i nauki. Ideę tę pięknie i przejrzyście wyklada w swym dziele. Myślą przewodnią jest ruch wirowy wokół punktu równowagi, czyli analiza zachowań cyklicznych typowa dla przyrody i ekonomii.

Książka nie ma charakteru historycznego, ale Autor nasycił ją historią nauki. Szczególnie cenne są anegdoty swoiste, które ożywiają lekturę i wnoszą głębsze zrozumienie problemu. Czytelnika ucieszy niewątpliwie opowieść o problemie trzech ciał, informacja o różnorodnych definicjach chaosu, opowieść hydrologiczna o wykładniku Hursta, definicja stabilności w sensie Lapunowa i temu podobne. Aneks jest jakby oddzielną książeczką bardzo ułatwiającą lekturę części zasadniczej. Stephen Hawking do trzech największych odkryć XX wieku zalicza chaos obok twierdzenia Goedela o niezupełności arytmetyki liczb naturalnych i zasady nieoznaczoności Heisenberga. Te ciekawe informacje Autora uzupełnię pytaniem o trzy najważniejsze osiągnięcia matematyki w całej kulturalnej historii ludzkości? Zdaniem niektórych specjalistów są to: metoda aksjomatyczna Euklidesa, zapis pozycyjny liczb naturalnych oraz probabilistyka. Pod tą listą chętnie się podpisuję z jednym zastrzeżeniem: wymieniam probabilistykę na metodę wyczerpywania Archimedesesa. Od Archimedesesa wiedzie prosta droga do rachunku różniczkowego i teorii form różniczkowych, a bez tych rzeczy nie ma mowy o współczesnej cywilizacji. Dynamikę systemu opisują formy różniczkowe i równania różniczkowe. Najważniejsze równania fizyki matematycznej są równaniami liniowymi i to, co najwyżej drugiego rzędu. Pierwsza pochodna jest aproksymacją lokalną funkcją liniową, a druga – funkcją dwuliniową. Praktycznie ważne są więc formy liniowe, które w ekonomii reprezentują ceny oraz formy kwadratowe, które w ekonomii odzwierciedlają inwestycje. Probabilistyka zaliczana do największych osiągnięć matematyki

jest naturalnie odpowiednikiem chaosu w klasyfikacji Hawkinga. W aneksie wykładnik Lapunowa definiuje się metodą zwaną przez Anglików *very sophisticated*. Jeśli funkcja $f: R \rightarrow X$ jest trajektorią w przestrzeni metrycznej stanów X z metryką d , to wykładnik Lapunowa λ_f funkcji f określa się prostym wzorem

$$\lambda_f = \limsup t^{-1} \ln d(a, f(t)), \text{ gdy czas } t \text{ rośnie nieograniczenie,}$$

gdzie R jest naturalnie zbiorem liczb rzeczywistych, natomiast a dowolnym stanem w przestrzeni X , a gdy X jest przestrzenią liniową, wtedy zwykle $a = 0$. Dla funkcji będącej rozwiązaniem wymienionego wyżej równania Malthusa – pomijając rozwiązanie trywialne równe stałe 0 – wykładnik Lapunowa jest równy liczbie r . Przy okazji zacytuję Autora (str. 108). *Należy zauważyć, że wszystkie badania numeryczne mają charakter przybliżony [...]. Z oczywistych powodów nie jesteśmy w stanie zbliżyć się do granicy [...]. Powoduje to, że nigdy nie będziemy pewni, czy znak wykładnika Lapunowa nie zmieni się po wykonaniu dostatecznie dużej liczby iteracji.* Z wypowiedzi tej wynika naturalny wniosek: badania asymptotyczne w ekonomii mają jedynie charakter teoretyczny, a wyciąganie z nich wniosków praktycznych stoi pod wielkim znakiem zapytania. Ten znak zapytania może powiększyć się, bowiem w innym miejscu Autor pisze: *W dynamice nieliniowej, stanowiącej sumę teorii chaosu deterministycznego i teorii złożoności dużą rolę odgrywają obliczenia* (s. 9).

W pracy są przynajmniej dwa rodzaje chaosu: chaos wśród modeli zależnych od parametrów i chaos trajektorii zależny od stanu początkowego. Autor operuje zbiorami miary 0 w sensie Lebesgue'a. Czy nie lepiej używać uniwersalnego pojęcia zbioru generycznego wprowadzonego w teorii katastrof, a obowiązującego w dowolnej topologicznej przestrzeni stanów? Zbiór generyczny, to otwarty i gęsty w całej przestrzeni podzbiór. Naturalnie dopełnienie zbioru miary zero nie zawsze jest zbiorem generycznym; w sytuacjach rzeczywistych zwykle bywa. Zbiór liczb wymiernych jest miary zero, ale jego dopełnienie – liczby niewymierne – nie tworzą zbioru generycznego. Istnieje swego rodzaju sprzeczność pomiędzy próbą uwolnienia się od chaosu deterministycznego a realnym wzrostem ekonomicznym. Rozwój układu cybernetycznego określonego wzorem $x_{n+1} = Ax_n$, gdzie wektory są koszykami dóbr, a A jest macierzą procesu technologicznego, zależy od wartości własnych macierzy A . Gdy wartości własne macierzy A są wszystkie mniejsze od 1, nawet dodatnie nie mówiąc o ujemnych, wtedy układ cybernetyczny zdefiniowany powyższym układem dynamicznym cofa się – nie ma rozwoju jest regres. Układ taki rozwija się jeśli maksymalna wartość własna jest silnie większa od 1. Wykładnik Lapunowa jest równy tej największej wartości własnej, mamy więc chaos, ale mamy jednocześnie rozwój; gdy nie ma chaosu jest stagnacja, układ dąży do samounicestwienia. Tak więc powstał paradoks: rozwój – więc chaos, zastój – więc stabilność i spokój. Trzeba wybierać i wszyscy wybierają chaos. Nie ma na to rady. Stabilny socjalizm porzuciliśmy, a przyjęliśmy niestabilny – chaotyczny – kapitalizm. Opisana tu dynamika jest naturalnie liniowa.

Książka poświęcona chaosowi deterministycznemu w ekonomii jest niezwykle aktualną w dzisiejszej dobie i oczekiwaną przez rynek. Chaos deterministyczny jest – obrazowo mówiąc – przejawem złożoności systemu. Jak w sytuacjach kryzysowych: pożar, powódź, huragan, wojna, zahamować strach i zachowania nieracjonalne i zmusić układ do powrotu do stanu równowagi stabilnej? Panika i strach destabilizuje układ; opanowanie zachowań stadnych musi być drastyczne i jest niezbędne. Utrudnia to panujący dzisiaj paradygmat w ekonomii: niewidzialna ręka rynku rodzi właśnie chaos. Potrzebny jest więc nowy język i nowe metody. Książka stara się zapełnić lukę pomiędzy tym co wie nauka a potrzebami praktyki i dydaktyki. Znane są przypadki wygranej bitwy po okiełzaniu początkowego rozprężenia – paniki i ucieczki z pola walki. Z ekonomicznego punktu widzenia, a w szczególności pracy giełdy, temat ten jest poruszany w głośnych ostatnio książkach Sorosa i Greenspana. Znani ci działacze, raczej praktycy niż akademicy, nie podają gotowych rozwiązań, a głównie zwracają uwagę na potrzeby uregulowań prawnych porządkujących działalność inwestycyjną graczy giełdowych i banków. W przyrodzie obserwuje się dwa sterowania układami cybernetycznymi. Sterowanie dalekosiężne na wybrany cel i działanie zmierzające do utrzymania organizmu w zdrowiu, a systemu w stanie sprawności technicznej. Sterowaniem dalekosiężnym winien zajmować się rząd, a sterowanie krótkookresowe wymusza zasada równowagi. Tak w ekonomii jedną sprawą jest wyrabianie potrzeb i wzorów konsumpcji, a inną zaspokajanie tych potrzeb. Te dwa sterowania bardzo obrazowo widać w instytucji lotów kosmicznych. W ośrodku kontrolnym lotu jest sterowanie rakieta na określony cel, na przykład Księżyc lub Marsa, a na pokładzie rakiety jest urządzenie automatyczne sterujące stanem tej rakiety, tak by nie uległa samozniszczeniu i mogła osiągnąć cel dalekosiężny. Te dwa sterowania są wyraźnie od siebie oddzielone. Podobnie jest w ekonomii. To rozróżnienie sterowań jest podstawą podziału ekonomii na naukę makro i mikro. Rakieta jest całością: nawet gdy doleci do miejsca przeznaczenia, ale

na pokładzie jest źle, to zadanie nie zostanie zrealizowane. Nie ma zdrowego rozwoju ekonomii społecznej bez zdrowych przedsiębiorstw. Rząd wyznacza cele odległe i racjonalne, a rynek powoduje, że organizm społeczny – sprawny i w zdrowiu – dąży do tych dalekich przeznaczeń. Wchodząc nieco na teren filozofii warto się zastanowić nad pytaniem: w jakim stopniu człowiek określa swój los? Być może wszystko jest zdeterminowane przez Ananke – przeznaczenie, a z wielu możliwości nie wybieramy tego, co chcemy, tylko ktoś wybiera to, co jest konieczne. Chociaż mamy wolną wolę, realizujemy przeznaczenie.

Jak już powiedziałem książka robi dobre wrażenie. Jednakowoż nie jest idealna, bo nie ma rzeczy – stworzonych przez człowieka – bez wad i niedoskonałości. Po pierwsze, rozdziały są niejednakowej wielkości – niejednorodne. Rozdział IV jest hołdem złożonym Cournot. Ten krótki rozdział pasuje bardzo do wstępu, podobnie jak rozdział ostatni do aneksu. Przez takie zmiany redakcyjne książka stałaby się jednolita i spójna. Dobrze o tym pamiętać jeśli dojdzie do wznowień. Rozdziały V i VI można łatwo połączyć, wystarczy bowiem analizować hiperboliczne funkcje zysku producentów w postaci

$$f_i(x) = (x_1 + \dots + x_n)^{-1} x_i - c_i x_i,$$

gdzie $i = 1, \dots, n$ oraz $x \in R^n$; otrzymuje się w ten sposób ekonomiczny analogon n ciał, czyli rynek n producentów. Czym jest złożoność? Temu pojęciu Autor poświęca wiele miejsca i wysiłku, bo jest to niewątpliwie cecha charakterystyczna takiego układu cybernetycznego jak ekonomia społeczna. Osobiście uważam, że złożoność to rzecz niezwykle prosta – przyroda i natura jest oparta na prawach nieskomplikowanych. Złożoność jest rekurencją. Tak jak trudne i kręte schody na wysoką wieżę są utworzone z wieluset łatwych i prostych stopni. A więc rekurencja jest podstawą nie tylko biologii, ale także ekonomii: powtórzenia wielokrotnie według jakiejś prostej jednej wspólnej formuły. Złożoność jest w jakimś sensie naszą niewiedzą o tym, że świat jest prosty. Prawda jest jedna, ale teorii naukowych dużo. Fizycy od lat szukają jednolitej teorii opisującej wszystko; ciągle na próżno. Autor ma niewątpliwie słuszość, gdy twierdzi, że nauka jest jedna, a poszczególne dziedziny czy branże to swoista interpretacja ogólnej powszechnej teorii. Nauka jest więc strukturą, czyli jakby syntaksą, a konkretne dziedziny interpretacjami, czyli semantyką. W nauce nie ma więc prawdy, prawda jest w zastosowaniach nauki. Autor nadaje – myślę, że przesadne – znaczenie osiągnięciom fizyki. Fizycy od prawdy są równie daleko jak ekonomiści. Termodynamiczna interpretacja ekonomii jest niewątpliwie ciekawa; na tej drodze można odkryć wiele prawidłowości ekonomicznych. Analogia i izomorfizm jest podstawą nauki. Każda teoria jest jakimś pomiarem, a więc homomorfizmem, rzeczywistości w strukturę formalną. Odwzorowujemy obiekty realne w pojęcia abstrakcyjne. Nauka niewątpliwie upraszcza sprawy trudne i nie komplikuje tych prostych; taka jest istota homomorfizmu. Unikałbym – język Autora jest jak powiedziałem piękny – słów w rodzaju oksymoron, intermitencja *et cetera* bezpośrednio przyjętych z łaciny, greki lub angielskiego. Nie świadczą o uczoności, utrudniają lekturę, a najważniejsze – psują rytm języka polskiego. Wiele nierównoważnych definicji złożoności świadczy o młodości tego pojęcia i na klasykę potrzeba jeszcze poczekać. Proponuję w tej materii to, co wyżej napisałem: złożoność jest iteracją. Teoria chaosu pokazuje prostotę złożoności. Wszystko otrzymasz wielokrotnie iterując jedną funkcję. Za chaosem ekonomicznym niewątpliwie stoi porządek i harmonia fal Elliotta. Istotą jest – powtarzam już po wtóre – zasada równowagi: wirowanie stanu równowagi chwilowej wokół idealnego punktu, do którego zmierza układ.

Autor książki wnosi osobiste zaangażowanie i intuicyjne odczucia do swego naukowego dzieła. Poglądy Autora są mi bliskie i z nimi absolutnie się zgadzam. Na postęp nauki mają także często zasadniczy wpływ pomyłki i błędy uczonych. Jest na to wiele przykładów, a wymienię tylko Elliotta i Poincarego. Elliott wykrył spiralę, ale nie umieścił jej tam, gdzie trzeba. Bez Elliotta nie byłoby teorii wirów ekonomicznych i spiralnego rozwoju objaśniającego cykliczność. Świat jest prosty, rządzonej jednym prawem – prawem równowagi: akcja jest reakcją. Chociaż znamy zasady ogólne, praktycznie trudno przewidzieć tor komety. Przy okazji omawiania chaosu Autor ładnie objaśnia głęboki i przez długie lata zapomniany twierdzenie Szarkowskiego o cykliczności. Prawa nauki mają tylko przybliżony charakter. Nigdy nie ma pewności realizacji określonego stanu i to zarówno w ekonomii, jak i fizyce. Takie pojęcia jak wykładnik Hursta, wykładnik Lapunowa i inne pokrewne mają charakter techniczny – ułatwiają podejmowanie decyzji jak większość narzędzi formalnych. Świat istotnie jest lokalnie liniowy i gładki; gładkość oznacza istnienie pochodnej. Pochodna jest optymalną lokalną aproksymacją najbardziej brzydkiej nieliniowej funkcji. Słowo *najbardziej* nic tu nie znaczy. Ma sens czysto ekspresyjny. Wiara w liniowość natury jest stwierdzeniem w moim odczuciu bezzasadnym. Nie ma takiej wiary. Świat wydaje się gładki, gdy patrzymy z odpowiedniego dystansu lub całkowicie porwany,

gdy patrzymy w mikroskali. Natura jest taka jaka jest. Owszem można powiedzieć, że istnieje powszechna wiara w różniczkowalność, czyli gładkość. Już sama ciągłość pozwala na lokalną aproksymację liniową, ale nie optymalną w sensie styczności. Liniowość obowiązuje w przestrzeniach liniowych. W ekonomii nie ma liniowości z formalnego punktu widzenia, bo nie ma dowolnie dużo rozporządzalnych zasobów; zamiast addytywności jest podaddytywność

$$f(x + y) \leq f(x) + f(y)$$

kosztów lub nadaddytywność

$$f(x + y) \geq f(x) + f(y)$$

efektów – synergia. Współpraca zawsze jest pożądana – współpraca zwiększa produkt globalny; wszyscy są wygrani. Świat jest jaki jest, czyli nieliniowy. Wzrost naturalny jest, jak wiemy, wykładniczy; ale taki wzrost może być krótkookresowy z powodu braku zasobów. Liniowość nie ma wiele wspólnego z determinizmem. Prawo nauki i warunki początkowe definiują układ stabilny, modele zaś mogą generować stabilność lub jej brak – zależy to od rodzaju modelu.

Autor niewątpliwie pokazuje, że ekonomia jest nauką barwną. Ekonomia jest także sztuką. Recenzowane dzieło jest wartościowe, godne polecenia. Autor jest przecież już znanym i cenionym w świecie naukowym twórcą. Wiele jest rzeczy ładnych i wartościowych w tej niezwykle książce; jest to rzecz pożyteczna i piękna chociaż daleka od doskonałości. Pobudza do dyskusji, inspiruje twórczo; po jej przeczytaniu ma się ochotę do zaangażowania w te niezwykle badania, by albo wesprzeć Autora, albo z nim polemizować. Doskonałość jest poza zasięgiem ludzkich możliwości, ideałów Platońskich nie ma w świecie materialnym. Ale nauka ma tam dążyć. Autor o tym wie i chce być blisko doskonałości.

SPRAWOZDANIA

JADWIGA SUCHECKA, MACIEJ JEWCAK

I MIĘDZYNARODOWA KONFERENCJA NAUKOWA EKONOMETRIA PRZESTRZENNA I REGIONALNE ANALIZY EKONOMICZNE

W dniach 22-23 czerwca 2010 roku odbyła się I Międzynarodowa Konferencja Naukowa nt.: „Ekonometria Przestrzenna i Regionalne Analizy Ekonomiczne”. Konferencja została zorganizowana przez Katedrę Ekonometrii Przestrzennej Uniwersytetu Łódzkiego. Obrady odbyły się w Centrum Szkoleniowo-Konferencyjnym Uniwersytetu Łódzkiego oraz na Wydziale Zarządzania UŁ.

Celem konferencji była prezentacja nowych teorii, metod i narzędzi badań empirycznych z zakresu ekonometrii przestrzennej, statystyki przestrzennej oraz regionalistyki, a także integracja środowiska naukowców zajmujących się teoretycznymi i praktycznymi analizami ekonomicznymi w ujęciu regionalnym.

W konferencji wzięło udział około 100 osób z wielu ośrodków naukowych z zagranicy i z Polski oraz Urzędów Statystycznych. W prezentacji swoich osiągnięć naukowych brali również udział studenci IV i V roku specjalności Modelowanie i Prognozowanie Regionalne, kierunek Gospodarka Przestrzenna, Uniwersytet Łódzki. Studenci, pomimo młodego stażu w prowadzeniu badań naukowych, wykazali się bardzo wysokim poziomem wiedzy i umiejętnością prowadzenia dyskusji naukowej. Sesja studencka zdobyła wysokie uznanie uczestników konferencji, a najlepsze pod względem naukowym referaty na wniosek Komisji Oceniającej zostały nagrodzone przez organizatorów. Ze względu na czynny udział licznej grupy studentów konferencja ta miała także charakter edukacyjny.

W trakcie dwudniowych obrad zostało wygłoszonych 41 referatów w 8 sesjach tematycznych i zaprezentowano 16 prac w sesji plakatowej.

Poniżej prezentujemy krótką charakterystykę referatów prezentowanych w kolejnych sesjach. Z treścią wygłoszonych referatów oraz zaprezentowanych prac będzie można zapoznać się po wydaniu przez Uniwersytet Łódzki trzytomowej monografii.

Uwzględniając duże zainteresowanie problematyką konferencji organizatorzy planują następne spotkanie w 2012 roku.

Honorowy patronat objął profesor Jean Henri Paul Paelinck z Uniwersytetu George’a Masona, a wykłady inauguracyjne wygłosili wybitni specjaliści z zakresu ekonometrii i statystyki przestrzennej: profesor Raymond Florax z Uniwersytetu Purdue i profesor Anil K. Bera z Uniwersytetu Illinois.

Prof. Anil K. Bera (Uniwersytet Illinois) wygłosił wykład nt.: *Spatial Analysis: From the beginning to the Frontiers of Spatial Econometrics*. Problematykę wykładu skoncentrował wokół najistotniejszych problemów związanych z zależnościami przestrzennymi charakteryzującymi procesy/zjawiska ekonomiczne i społeczne. Prof. Bera zaprezentował historię analiz przestrzennych, poczynawszy od analizy przestrzenno-czasowej rozprzestrzeniania się epidemii cholery w Londynie w latach 50. XIX w. W dalszej części wykładu, profesor Bera omówił sylwetki osób uznawanych za prekursorów przestrzennych analiz, skupiając się na również na pracach Grangera i Morana. Wykład kończyła prezentacja problemów zależności nieliniowych w analizach przestrzennych wraz z przykładem zastosowań do analizy kształtowania się cen nieruchomości w Bostonie.

Prof. Raymond Florax w swoim wykładzie nt.: *Spatial Heterogeneity or Dependence: That's the Question* skoncentrował uwagę na istotności źródeł przestrzennej heterogeniczności i sposobach jej analizy. Profesor R. Florax wskazując na konieczność uwzględnienia i odróżnienia reżimów przestrzennych od interakcji pomiędzy obiektami, pokazał możliwości zastosowań geograficznie ważonej regresji (GWR) oraz przestrzennie wygładzonych modeli autoregresji STAR. Z punktu widzenia analiz przestrzennych ważne są również metody testowania istotności źródeł heterogeniczności przestrzennej np. wykorzystujące mnożniki Lagrange'a oraz eksperymenty Monte Carlo. Na zakończenie wykładu zaprezentowane zostały wyniki autorских badań dotyczące określenia poziomu bogactwa USA dla lat 1990-2007 z zastosowaniem modelu STAR. Pozostałe wystąpienia uczestników konferencji odbywały się w kolejnych sesjach równoległych.

W sesji I: **Wzrost regionalny** zostało przedstawionych 5 referatów.

Prof. Christian Le Bas (LEFI, Institut des Science de L'Homme, Uniwersytet Lyon 2, Francja), *Spatial distribution and inter regional mobility of French prolific investors. An empirical analysis*. Prezentacja dotyczyła empirycznej analizy rozkładu przestrzennej i międzyregionalnej mobilności francuskich „wynalazców”. Celem tych analiz było wskazanie geograficznej lokalizacji, ponad regionalnej mobilności oraz związków między indywidualnymi decyzjami wynalazców (skala mobilności), a ich produktywnością (liczba patentów stosowanych w czasie trwania tych ruchów).

Dr Dorota Górecka (Uniwersytet im. Mikołaja Kopernika w Toruniu) w referacie pt.: *Analiza przestrzenna innowacyjności polskich regionów* przedstawiła najważniejsze problemy dotyczące analizy przestrzennej innowacyjności polskich regionów zwracając uwagę na rosnące znaczenie innowacyjności w rozwoju nie tylko przedsiębiorstw, ale i regionów. W podsumowaniu rozważań potwierdziła tezę, że innowacyjność gospodarek stanowi jedno z ważnych źródeł ich przewagi konkurencyjnej.

Prof. Władysław Welfe, dr Piotr Karp (Uniwersytet Łódzki) w referacie: *The structural model of the Lodz region – scenarios of regional development* zaprezentowali scenariusze rozwoju regionu łódzkiego. Podstawę analiz stanowił strukturalny makroekonomiczny model dla regionu łódzkiego LORIS wraz z zaadaptowanym modelem gospodarki Polski W8-LW.

Prof. Jadwiga Suchecka, Edyta Łaszkiewicz (Uniwersytet Łódzki): *Odległość geograficzna i ekonomiczna a zmiany powiązań europejskich rynków giełdowych w okresie kryzysu 2007-2009*. Autorki w swoim wystąpieniu podkreśliły rosnące znaczenie zależności przestrzennych w analizach giełdowych. Przedstawiły wyniki weryfikacji hipotez dotyczących związków przestrzennych pomiędzy powiązаныmi giełdami, wykorzystując w tym celu model DCC GARCH oraz specyficzne narzędzia analiz przestrzennych.

Dr Grzegorz Micek (Uniwersytet Jagielloński): *Spatial problems of estimating multiplier effects in the local scale*. Tematem wystąpienia były przestrzenne problemy szacowania efektów mnożnikowych w skali lokalnej. Autor wskazał, że studia lokalnych efektów mnożnikowych funkcjonowania dużych przedsiębiorstw, wybranych sektorów gospodarki czy firm zlokalizowanych w specjalnej strefie ekonomicznej należą w naszym kraju do rzadkości. Sytuacja ta wynika głównie z konieczności stosowania odmiennego podejścia od klasycznej analizy nakładów-wyników.

Drugą sesją równoległą była sesja zatytułowana **Ekonomia zdrowia**, w której zaprezentowano 4 referaty.

Prof. Anatoly Pilyavsky (Lviv Academy of Commerce, Ukraina) w referacie: *Efficiency differences among hospitals and surgical departments in Ukraine* zaprezentował zastosowanie nieparametrycznej metody DEA do analizy przestrzennego zróżnicowania efektywności nakładów na szpitalną opiekę chirurgiczną w wyróżnionych regionach Ukrainy. Wyniki badań świadczą o przestrzennych nierównościach w nakładach i wynikach tego typu świadczeń zdrowotnych.

Dr Iwona Laskowska (Uniwersytet Łódzki), *Kapitał zdrowotny a rozwój regionalny. Podejście panelowe*. Autorka w swoim referacie zaprezentowała zastosowanie dynamicznego modelu panelowego do analiz kapitału zdrowotnego w powiązaniu z rozwojem regionalnym, odwołując się również do prowadzonych badań nad wzrostem gospodarczym oraz zaobserwowanym rosnącym zainteresowaniem potencjalnymi

korzyściami ekonomicznymi wynikającymi z poprawy stanu zdrowia populacji. Takie podejście pozwoliło na określenie związku między zróżnicowaniem wskaźników zdrowotnych i różnicami w poziomie PKB.

Mgr Anna Pysik, mgr Sebastian Twaróg (Uniwersytet Ekonomiczny w Krakowie) w referacie: *Badanie autokorelacji przestrzennej krwiodawstwa w Polsce* prezentowali wyniki badań własnych dotyczących występowania zależności przestrzennej w kształtowaniu krwiodawstwa w Polsce. Z przeprowadzonych analiz wynika, że występuje autokorelacja przestrzenna badanego zjawiska, a zastosowanie narzędzi statystyki przestrzennej umożliwiło wyróżnienie głównych centrów rozwoju krwiodawstwa w kraju.

Dr Paweł Kretowicz (Uniwersytet Jagielloński) w referacie *The utilisation of canonical analysis in explaining factors influencing hospitalization rates on a regional scale* zaprezentował możliwości zastosowania analizy kanonicznej do oceny natężenia i przestrzennego zróżnicowania przyczyn wpływających na różne wartości współczynników hospitalizacji ludności w regionie. W podsumowaniu swojego wystąpienia wskazał na wady i zalety analizy kanonicznej dla wyjaśniania przestrzennych nierówności w opiece szpitalnej, a także wskazał na możliwości pogłębienia analizy zróżnicowań poziomu zdrowia na poziomie regionalnym.

Kolejne dwie sesje dotyczyły analiz rynków pracy oraz metod analizy danych przestrzennych i panelowych.

W sesji **Rynki pracy** uczestnicy konferencji zaprezentowali 4 referaty związane z tematem tej sesji.

Dr Ewa Kusideł (Uniwersytet Łódzki), *Convergence on local Labor markets in Poland*. Celem wystąpienia było zweryfikowanie hipotezy o występowaniu konwergencji na lokalnych rynkach pracy w Polsce. Wnioskowanie wyprowadzono na podstawie standardowych narzędzi stosowanych do weryfikacji konwergencji typu *beta* i *sigma* oraz danych statystycznych z lat 1999-2008 pochodzących z Badania Aktywności Ekonomicznej Ludności.

Dr Artur Gajdos (Uniwersytet Łódzki) w referacie: *Przestrzenno-czasowa analiza struktury pracujących według wielkich grup zawodowych* prezentował wyniki badań dotyczących dynamiki i struktury pracujących według wielkich grup zawodowych w Polsce na poziomie wojewódzkim. Podjęta analiza zróżnicowania przestrzennego jakości kapitału ludzkiego, w ocenie Autora, stanowi główny element zainteresowania polityki spójności.

Dr Stanisław Matusik (Akademia Wychowania Fizycznego w Krakowie), *Ekonometryczne modele zmian stopy bezrobocia rejestrowanego w Polsce oraz województwie małopolskim w latach 1997-2009* zaprezentował wyniki analiz porównawczych dotyczących kształtowania się stopy bezrobocia rejestrowanego w Polsce oraz w powiatach i gminach województwa małopolskiego (dla roku 2009 w odniesieniu do 2002 r.).

Dr Iwona Müller-Frączek, dr Michał Bernard Pietrzak (Uniwersytet im. Mikołaja Kopernika w Toruniu): *Analiza stopy bezrobocia w Polsce z wykorzystaniem przestrzennego modelu MESS*. Przedstawili analizy stopy bezrobocia w Polsce z wykorzystaniem przestrzennego modelu MESS. Uzyskane wyniki badań potwierdziły jednoznacznie potrzebę uwzględniania zależności przestrzennych w analizie tak specyficznego procesu ekonomicznego, jakim jest bezrobocie.

W sesji **Metody analizy danych przestrzennych i panelowych** zaprezentowano 7 referatów.

Prof. Elżbieta Szulc, dr Iwona Müller-Frączek, dr Michał Bernard Pietrzak (Uniwersytet im. Mikołaja Kopernika w Toruniu), *Modelowanie zależności między ekonomicznymi procesami przestrzennymi a poziomem agregacji danych*. Autorzy w swoim wystąpieniu skoncentrowali się na wykazaniu określonych konsekwencji dla możliwości odkrywania różnych własności i struktur przestrzennych w sytuacji, gdy analizy statystyczno-ekonometryczne zjawisk przestrzennych są przeprowadzane przy wykorzystaniu danych o różnym poziomie agregacji przestrzennej. W rezultacie tego typu analiz pojawia się problem „naiwnej” interpretacji modeli przestrzennych.

Mgr inż. Monika Dyduch (Akademia Ekonomiczna im. K. Adamieckiego w Katowicach), *Przestrzenno-czasowa analiza handlu zagranicznego w Polsce*. Wystąpienie było poświęcone analizie prze-

strzenno-czasowej handlu zagranicznego Polski. Badanie było przeprowadzone z wykorzystaniem danych pochodzących z lat 2000-2010. Wyniki analiz obejmowały nie tylko prezentację struktury handlu zagranicznego Polski, ale także prognozy dla analizowanego zjawiska z uwzględnieniem danych o wysokiej częstotliwości otrzymanej na podstawie modelu, który był zintegrowany z transformatą falkową, siecią neuronową i algorytmem genetycznym.

Mgr Jan Acedański (Akademia Ekonomiczna im. K. Adamieckiego w Katowicach), *A bias of structure forecasts from Markov chain models*. Autor skupił swoją uwagę na obciążeniu prognoz struktury uzyskanych przy pomocy łańcuchów Markowa. Celem wystąpienia była ocena obciążenia prognoz uzyskanych tą metodą i ilustracja tego problemu na przykładzie prognozowania podziału produkcji globalnej pomiędzy regiony Polski.

Dr Tomasz Żądło (Akademia Ekonomiczna im. K. Adamieckiego w Katowicach), *O pewnych estymatorach kalibrowanych wartości globalnej w podpopulacji w oparciu o dane przekrojowo-czasowe*. Punktem wyjścia prezentacji było stwierdzenie, iż modelowanie profili wielookresowych wymaga przyjęcia różnych założeń. W swoich badaniach Autor przyjął założenie, że zarówno populacja, jak i przynależność elementów populacji do podpopulacji, może zmieniać się w czasie. W badaniu symulacyjnym dokładność estymatorów kalibrowanych została porównana z dokładnością estymatora Horvitz-Thompsona oraz wyprowadzonego dla rozważanego modelu empirycznego najlepszego liniowego nieobciążonego predyktora.

Dr Wacław Sklinsmont (Wyższa Szkoła Informatyki i Ekonomii Towarzystwa Wiedzy Powszechnej w Olsztynie), *Macierze wielowskaźnikowe w naukach ekonomiczno-społecznych*. Autor rozpoczął swoje wystąpienie od podkreślenia, że obecny zakres zastosowań metod ekonometrii przestrzennej jest ciągle niewystarczający, aby podejmować kompleksowe badania procesów rozwoju społeczno-gospodarczego. Jednym z możliwych sposobów przezwyciężenia tego problemu jest wykorzystanie macierzy wielowskaźnikowych, które w sposób znaczący mogą przyspieszyć postęp w kompleksowych badaniach procesów społeczno-gospodarczych. W prezentacji podano definicję macierzy wielowskaźnikowych, działania na nich ze szczególnym akcentem na ich iloczyn, którego definicja obejmuje oprócz 5 znanych i stosowanych w praktyce, również 8 innych definicji mniej znanych.

Dr Second Bwanakare (Wyższa Szkoła Informatyki i Zarządzania w Rzeszowie), w referacie: *Non-parametric entropy approach for the balancing of Polish social accounting matrix (SAM) 2005*. I przedstawił podejście nieparametryczne do procesów samoradnych w równoważeniu polskich rachunków społecznych w macierzy SAM 2005. Celem prezentacji było określenie niezagregowanej środowiskowej macierzy SAM, która umożliwiła włączenie zjawisk ekologicznych w okrężny obieg gospodarki narodowej.

Dr Tomasz Kossowski, dr Jan Hauke (Uniwersytet im. Adama Mickiewicza w Poznaniu). Swoje wystąpienie poświęcili prezentacji metody obliczania logarytmu jacobianu transformacji zmiennych w modelach przestrzennych oraz kilku praktycznych przykładów odnoszących się do zastosowań w modelach interakcji przestrzennych.

Sesja V poświęcona była problematyce **Zrównoważonego rozwoju i konwergencji regionalnej**.

Prof. Jadwiga Suchecka, mgr Elżbieta Antczak (Uniwersytet Łódzki), w referacie: *Application of spatial panel autoregressive models to sustainable development level assessment in European countries* zaprezentowały możliwe aplikacje modeli panelowych z efektami przestrzennymi do oceny poziomu realizacji założeń zrównoważonego rozwoju w 32 krajach Europy. W prezentacji zwrócono również uwagę na metody estymacji tego typu modeli.

Dr Monika Stanny (Instytut Rozwoju Wsi i Rolnictwa PAN), w referacie pt.: *Przestrzenna analiza poziomu komponentów rozwoju zrównoważonego i ocena stopnia ich zrównoważenia na obszarach wiejskich Zielonych płuc Polski* zwróciła szczególną uwagę na sposób konstrukcji syntetycznego wskaźnika oceny komponentów rozwoju zrównoważonego i jego zastosowaniu do zbadania stopnia ich zrównoważenia.

Dr Barbara Dańska-Borsiak (Uniwersytet Łódzki), *Covergence or divergence of Polish voivodeships? Analysis based on dynamic panel data model*. Autorka zaprezentowała wyniki swoich badań dotyczące zastosowania modeli dynamicznych szacowanych na podstawie danych panelowych. Szczególna uwaga została poświęcona alternatywnym wersjom Uogólnionej Metody Momentów (GMM): metodzie pierwszych różnic Arellano i Bonda oraz metodzie systemowej Blundella i Bonda.

Dr Beata Bal-Domańska (Uniwersytet Ekonomiczny we Wrocławiu), w referacie *Ekonometryczna identyfikacja β konwergencji regionów szerebla NUTS-2 państw Unii Europejskiej* uwzględniła w analizie czynniki wynikające z konstrukcji rozszerzonego modelu Solowa (model Mankiwa-Romera-Weila), a do oceny procesów konwergencji wykorzystała wyniki oszacowań otrzymane na podstawie panelu danych z wykorzystaniem estymatora systemowego uogólnionej metody momentów.

Mgr Adrian Burdziak (Uniwersytet Łódzki), w referacie: *Empiryczna weryfikacja efektów aglomeracji w polskich podregionach przy wykorzystaniu metod panelowych* zaprezentował wyniki badań empirycznych dotyczących zastosowania modeli panelowych do weryfikacji efektów aglomeracji w polskich podregionach. W podsumowaniu Autor stwierdził, że gospodarka Polski odznacza się rosnącym trendem zróżnicowania międzyregionalnego, a zróżnicowanie efektywności ekonomicznej podmiotów gospodarczych może być objaśnione poprzez identyfikację efektów aglomeracji.

W sesji VI **Rynki nieruchomości i finansowe, analizy sektorowe**, przedstawiono następujące referaty.

Dr Mariola Chrzanowska, dr inż. Krzysztof Kompa (Szkoła Główna Gospodarstwa Wiejskiego w Warszawie), w referacie pt.: *Zastosowanie wielowymiarowej analizy porównawczej do klasyfikacji giełd europejskich* dokonali porównania europejskich giełd-przedsiębiorstw za pomocą wybranych klasycznych miar statystycznych. Wynikiem tych badań był utworzony ranking giełd przeprowadzony z uwzględnieniem wskaźnika względnego poziomu rozwoju (BZW), miary atrakcyjności inwestycji (TMAI), metody Warda oraz metody *k*-średnich.

Dr Artur Lipieta, dr hab. Barbara Pawełek (Uniwersytet Ekonomiczny w Krakowie), *Badanie przestrzenno-czasowego zróżnicowania redystrybucji dochodu narodowego brutto Unii Europejskiej w latach 2004-2008*. Autorzy zaprezentowali wyniki analizy statystycznej dotyczącej przestrzenno-czasowego zróżnicowania redystrybucji DNB Unii Europejskiej z zastosowaniem metod statystycznej analizy danych, współczynników potencjału przestrzennego, modelu dla danych panelowych oraz analizy korespondencji.

Prof. Barbara Podolec, dr Paweł Ulman, mgr Agnieszka Wałęga (Uniwersytet Ekonomiczny w Krakowie), *Próba oceny zróżnicowania spożycia artykułów żywnościowych w Polsce na podstawie wyników badań budżetów gospodarstw domowych*. Autorzy dokonali oceny zróżnicowania spożycia artykułów żywnościowych w Polsce na podstawie wyników badań budżetów gospodarstw domowych, koncentrując się na analizie zgodności spożycia zaobserwowanego w gospodarstwie domowym z wzorcem. Badanie przeprowadzono w ujęciu regionalnym z uwzględnieniem klasy miejscowości zamieszkania głowy gospodarstwa domowego.

Sesja VII poświęcona była **Zastosowaniu metod i modeli ekonometrii przestrzennej**.

Dr Tomasz Brodzicki, dr Dorota Ciołek (Uniwersytet Gdański), *Przestrzenno-czasowa analiza determinant lokalizacji przemysłu przetwórczego i usług rynkowych w polskich regionach*. Autorzy prezentując wyniki swoich badań z zastosowaniem metod analiz przestrzennych podkreślili, że sektory przemysłu o wyższym poziomie zaawansowania technologicznego, jak i sektory usług o wyższej intensywności wiedzy, lokują się w regionach z rozbudowaną bazą badawczo-rozwojową.

Prof. Krystian Heffner, mgr Piotr Gibas (Akademia Ekonomiczna im. K. Adamieckiego w Katowicach), w referacie: *Metody taksonomiczne w polityce regionalnej. Analiza porównawcza interpretacji wyników* dokonali autorskiej próby oceny poszczególnych metod taksonomicznych z punktu widzenia przydatności ich wyników dla praktycznych aspektów badań regionalnych. Wskazali również na mocne

i słabe strony omawianych metod, ze szczególnym uwzględnieniem poprawności i zakresu wnioskowania, przeprowadzonego na ich podstawie.

Dr Krzysztof Janc, dr Dariusz Ilnicki, dr Maciej Kryza, dr Mariusz Szymanowski (Uniwersytet Wrocławski), *Cechy rozmieszczenia sklepów w przestrzeni miejskiej na przykładzie Wrocławia*. Celem wystąpienia było przedstawienie możliwości zastosowania wybranych metod analizy przestrzennej do określenia zależności pomiędzy liczbą ludności a rozmieszczeniem sklepów w przestrzeni miasta.

Mgr Przemysław Szczuciński (Państwowa Wyższa Szkoła Zawodowa w Gorzowie Wielkopolskim), *Modelowanie związków sektora MPS oraz inwestycji zagranicznych z zatrudnieniem i wzrostem gospodarczym w regionach*. Celem prezentowanych wyników badań była ocena zróżnicowania poziomu wzrostu regionalnego oraz określenie wpływu wskazanych czynników na to zróżnicowanie.

Dr Michał Bernard Pietrzak, mgr Emilia Elżbieta Rutkowska (Uniwersytet im. Mikołaja Kopernika w Toruniu), *Analiza czasu dojazdu do centrum miasta z wykorzystaniem modelu przestrzennego SAR*. W swoich badaniach Autorzy założyli istnienie teoretycznego miasta podzielonego na dzielnice. Przy tym założeniu zbadali czas dojazdu mieszkańców do centrum drogami miejskimi. Klasyfikacji dróg dokonano z uwzględnieniem ich cech charakterystycznych dla dobra publicznego. Do rozwiązania problemu został wykorzystany model przestrzenny SAR uwzględniający zależności przestrzenne.

Prof. Elżbieta Sojka (Akademia Ekonomiczna im. K. Adamieckiego w Katowicach), w referacie: *Analiza porównawcza struktur i procesów ludnościowych w wybranych krajach UE z wykorzystaniem metod taksonomicznych* Autorka zaprezentowała porównanie zmian w strukturze ludności według wieku oraz zmian w podstawowych procesach demograficznych (płodność, dzietność, umieralność) w wybranych krajach Unii Europejskiej. Wykorzystując metody taksonomiczne zbadano podobieństwo struktur ludnościowych oraz podobieństwo krajów pod względem analizowanych procesów demograficznych.

Sesja VIII **Studencka** poświęcona była prezentacji wyników badań studentów w ramach specjalności „**Modelowanie i prognozowanie regionalne**” prowadzonej przez Katedrę Ekonometrii Przestrzennej UŁ na kierunku Gospodarka Przestrzenna.

P. Jasiński, M. Gołębiowski *Przestrzenne zróżnicowanie współczynnika aktywności zawodowej i stopy bezrobocia w krajach UE25*

R. Jaworska *Analiza zrównoważonego rozwoju województwa łódzkiego na tle Polski*

E. Łaskiewicz *Zastosowanie modelu VAR-GARCH w badaniu efektu zarażania holdingów finansowych*

E. Modranka *Rozwój gospodarczy województw w świetle teorii trzech sektorów – analiza prognoz ex-post modelu ewaluacyjnego EUImpactMod II*

P. Pabiańska *Czynniki ekonomiczne w analizie zjawiska migracji pomiędzy województwami w Polsce*

P. Strożek, R. Sztarnel *Działalność innowacyjna w Polsce w ujęciu strukturalno-geograficznym*

K. Trzeźwińska, E. Żmurkow, A. Gajdos *Jakość zasobów pracy i innowacyjność regionalna w Polsce*

W sesji plakatowej zaprezentowano:

M. Biernacki, W. Ejsmond (Uniwersytet Ekonomiczny we Wrocławiu), *Efektywność kształcenia we wrocławskich liceach*

A. Bukowska-Piastrzyńska (Uniwersytet Łódzki), *Kierunki zmian w dostępności ochrony zdrowia w woj. łódzkim u progu XXI w.*

E.K. Dłubakowska-Puzio, M. Karpuk, K. Puzio (Politechnika Koszalińska), *Badanie stopnia zadowolenia turystów pobytem w hotelach na podstawie modelu ryzyka turystycznego*

K. Lewandowska-Gwarda (Uniwersytet Łódzki), *Multi-equation spatial econometric models – theory and applications*

- M. Jewczak (Uniwersytet Łódzki), *Sektor opieki zdrowotnej w Polsce w analizach strukturalno-geograficznych*
- K. Świetlik (Politechnika Gdańska), *Model cen na rynku pierwotnym i wtórnym mieszkań w wybranych miastach Polski w latach 2005-2010*
- M. Sasin (Politechnika Koszalińska), *Główne determinanty migracji stałych w Polsce w latach 2003-2008*
- P. Kijko (Uniwersytet Warmińsko-Mazurski), *Analiza przydatności wybranych metod ekonometrycznych do prognozowania cen na wtórnym rynku mieszkaniowym*
- G. Przekota, M. Klonowska-Matynia (Politechnika Koszalińska), *Motywy podejmowania szkoleń przez bezrobotnych w województwie zachodniopomorskim*
- A. Przybylska-Mazur (Akademia Ekonomiczna w Katowicach), *Województwo śląskie na tle innych regionów Polski w aspekcie poziomu wskaźnika inflacji*
- A. Pysik (Akademia Ekonomiczna w Katowicach), *Przestrzenne zróżnicowanie współczynnika zgonu w Polsce w latach 1998-2008*
- J. Rój (Uniwersytet Ekonomiczny w Poznaniu), *Innowacje farmaceutyczne a wzrost ekonomiczny*
- D. Śliwicki (Urząd Statystyczny Oddział w Toruniu), *Ekonometryczny model wybranych charakterystyk rynku pracy*
- A. Sulima (Uniwersytet Jagielloński), *Analiza taksonomiczna przestrzennego zróżnicowania poziomu gospodarczego gmin w województwie małopolskim w roku 2007-2009*
- A. Wolny-Dominiak (Akademia Ekonomiczna w Katowicach), *Analiza stóp taryf w ubezpieczeniach komunikacyjnych z zastosowaniem mieszanych modeli liniowych*
- A. Żółtaszek (Uniwersytet Łódzki), *Spatio-temporal analysis of health care expenditures in OECD states*

SPIS TREŚCI

Artur Prędkie – Estymacja zbioru możliwości produkcyjnych w ramach formalnego modelu statystycznego	3
Czesław Domański – Uwagi o testach Jarque’a-Bera	19
Piotr Wdowiński – Polityka gospodarcza w modelu cen i kursu walutowego	27
Ewa Wędrowska – Wykorzystanie entropii Shannona i jej uogólnień do badania rozkładu prawdopodobieństwa zmiennej losowej dyskretnej	39
Zbigniew Świtalski – Równowagi na rynkach z dwustronnymi preferencjami	54
Ewa Drabik – Wykorzystanie reguł aukcyjnych do handlu energią w Polsce	70
Michał Łukowski – Zastosowanie modelu drzewa trójmianowego do wyceny opcji menedżerskich	88
Barbara Będowska-Sójska, Agata Kliber – Zmienność zrealizowana wobec modeli GARCH i modeli zmienności stochastycznej na polskim rynku kapitałowym	105
Anna Czapkiewicz, Iwona Skalna – Model CAPM oraz model Famy i Frencha na Warszawskiej Giełdzie Papierów Wartościowych	128
Piotr Tarka – Modele zmiennych ukrytych – rozważania nad pomiarem i generowaniem konstruktów w sferze wartości konsumentów	142

DYDAKTYKA I NAUKA

Mirosław Szreder – Losowe i nielosowe próby w badaniach statystycznych	168
Grzegorz Krzykowski – Filozoficzne koncepcje estymacji statystycznej	175

POLSCY STATYSTYCY I MATEMATYCY

Mirosław Krzyśko – Profesor Marek Fisz (1910-1963). W setną rocznicę urodzin	183
--	-----

RECENZJE

Antoni Smoluk – Recenzja książki: Aleksander Jakimowicz, <i>Źródła niestabilności struktur rynkowych</i>	191
--	-----

SPRAWOZDANIA

Jadwiga Suchecka, Maciej Jewczak – I Międzynarodowa Konferencja Naukowa Ekonometria Przestrzenna i Regionalne Analizy Ekonomiczne	196
---	-----

CONTENTS

Artur Prędkie – Estimation of Production Set Based on Formal Statistical Model	3
Czesław Domański – Some Remarks on Jarque-Bera Tests	19
Piotr Wdowiński – Economic Policy in the Model of Prices and Exchange Rate	27
Ewa Wędrowska – Application of Shannon's Entropy and its Generalizations for Studies on the Discrete Random Variable Probability Distribution	39
Zbigniew Świtalski – Equilibria in Markets with Two-Sided Preferences	54
Ewa Drabik – Application of Auction Rules on the Polish Power Exchange	70
Michał Łukowski – Pricing of Employee Stock Options Using Trinomial Tree	88
Barbara Będowska-Sójka, Agata Kliber – Realized Volatility Versus GARCH and Stochastic Volatility Models. The Evidence from the WIG20 Index and the EUR/PLN Foreign Exchange Market	105
Anna Czapkiewicz, Iwona Skalna – The CAPM and Fama-French Models in Poland ...	128
Piotr Tarka – Latent Variable Models – Issues on Measurement and Finding Exact Constructs in Customers' Values	142

TEACHING AND RESEARCH

Mirosław Szreder – Probability and Nonprobability Sampling in Statistical Surveys	168
Grzegorz Krzykowski – Philosophical Concepts of Statistical	175

POLISH STATISTICIANS AND MATHEMATICIANS

Mirosław Krzyśko – Professor Marek Fisz (1910-1963). On 100 Birthday Anniversary	183
--	-----

REVIEWS

Antoni Smoluk – Book review: Aleksander Jakimowicz, <i>Źródła niestabilności struktur rynkowych</i>	191
--	-----

REPORTS

Jadwiga Suchecka, Maciej Jewczak – I International Conference Spatial Econometrics and Regional Economic Analyses	196
--	-----