ZASTOSOWANIA METOD MATEMATYCZNYCH W EKONOMII I ZARZĄDZANIU

Studia Ekonomiczne

ZESZYTY NAUKOWE
WYDZIAŁOWE

UNIWERSYTETU EKONOMICZNEGO

W KATOWICACH

ZASTOSOWANIA METOD MATEMATYCZNYCH W EKONOMII I ZARZĄDZANIU

Redaktorzy naukowi
Jerzy Mika
Katarzyna Zeug-Żebro

Katowice 2013

Komitet Redakcyjny
Krystyna Lisiecka (przewodnicząca), Anna Lebda-Wyborna (sekretarz),
Florian Kuźnik, Maria Michałowska, Antoni Niederliński, Irena Pyka,
Stanisław Swadźba, Tadeusz Trzaskalik, Janusz Wywiał, Teresa Żabińska

Komitet Redakcyjny Wydziału Zarządzania
Janusz Wywiał (redaktor naczelny), Tomasz Żądło (sekretarz),
Alojzy Czech, Jacek Szołtysek, Teresa Żabińska

Rada Programowa
Lorenzo Fattorini, Mario Glowik, Miloš Král, Bronisław Micherda,
Zdeněk Mikoláš, Marian Noga, Gwo-Hsiu Tzeng

Recenzent
Janusz Łyko

Redaktor
Karolina Koluch

Skład
Krzysztof Słaboń

© Copyright by Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach 2013

ISBN 978-83-7875-089-5
ISSN 2083-8611

Wszelkie prawa zastrzeżone. Każda reprodukcja lub adaptacja całości
bądź części niniejszej publikacji, niezależnie od zastosowanej
techniki reprodukcji, wymaga pisemnej zgody Wydawcy

WYDAWNICTWO UNIWERSYTETU EKONOMICZNEGO W KATOWICACH
ul. 1 Maja 50, 40-287 Katowice, tel. 32 257-76-30, fax 32 257-76-43
www.wydawnictwo.ue.katowice.pl, e-mail: wydawnictwo@ue.katowice.pl

SPIS TREŚCI

Anna Czopek
ANALIZA PORÓWNAWCZA EFEKTYWNOŚCI METOD REDUKCJI ZMIENNYCH – ANALIZA SKŁADOWYCH GŁÓWNYCH I ANALIZA CZYNNIKOWA 7
Summary 23

Katarzyna Jakowska-Suwalska
WIELOKRYTERIALNY, NIELINIOWY MODEL WIELKOŚCI ZAMÓWIENIA MATERIAŁÓW DLA KOPALNI WĘGLA KAMIENNEGO..... 24
Summary 34

Anna Janiga-Ćmiel
WYZNACZENIE OKRESU RÓWNOWAGI I STABILIZACJI DŁUGOOKRESOWEJ 35
Summary 44

Adrianna Mastalerz-Kodzis
TEORIA FAL ELLIOTTA A TEORIA FRAKTALI – PODOBIEŃSTWA I RÓŻNICE W PODEJŚCIU DO MODELOWANIA SZEREGÓW ORAZ OPISU ZACHOWAŃ INWESTORA..... 45
Summary 55

Monika Miśkiewicz-Nawrocka
MODELE EKONOMICZNE Z DYNAMIKĄ CHAOTYCZNĄ..... 56
Summary 66

Ewa Pośpiech
PODPISY NIEZAPRZECZALNE – PODPISY Z DODATKOWĄ FUNKCJONALNOŚCIĄ..... 67
Summary 79

Zygmunt Przybycin
DWUMIANOWY MODEL WYCENY OPCJI W WARUNKACH ROZMYTYCH INFORMACJI..... 80
Summary 90

Adam R. Szromek
POMIAR FUNKCJI TURYSTYCZNEJ OBSZARÓW ZA POMOCĄ WSKAŹNIKÓW FUNKCJI TURYSTYCZNEJ NA PRZYKŁADZIE OBSZARÓW PAŃSTW EUROPEJSKICH 91
Summary 103

Joanna Trzęsiok
OCENA WPŁYWU ZMIENNYCH OBJAŚNIAJĄCYCH NA ZMIENNĄ ZALEŻNĄ W METODZIE RZUTOWANIA PPR 104
Summary 114

Michał Trzęsiok
SYMULACYJNA OCENA JAKOŚCI ZAGREGOWANYCH MODELI ZBUDOWANYCH METODĄ WEKTORÓW NOŚNYCH 115
Summary 126

Łukasz Wachstiel
PORÓWNANIE METOD ROZMYTEGO I PROBABILISTYCZNEGO MODELOWANIA ZJAWISKA NA PRZYKŁADZIE OCENY RYZYKA USŁUG INFORMATYCZNYCH 127
Summary 135

Maciej Wolny
ASPEKT SYTUACJI STATUS QUO WE WSPOMAGANIU WIELOKRYTERIALNEGO WYBORU BAZUJĄCEGO NA TEORII GIER 136
Summary 147

Maciej Wolny
WSPOMAGANIE KOOPERACJI Z WYKORZYSTANIEM TEORII GIER I ANALIZY WIELOKRYTERIALNEJ 148
Summary 159

Katarzyna Zeug-Żebro
WPŁYW REDUKCJI POZIOMU SZUMU LOSOWEGO METODĄ NAJBLIŻSZYCH SĄSIADÓW NA WARTOŚĆ NAJWIĘKSZEGO WYKŁADNIKA LAPUNOWA 160
Summary 168

Anna Czopek

ANALIZA PORÓWNAWCZA EFEKTYWNOŚCI METOD REDUKCJI ZMIENNYCH – ANALIZA SKŁADOWYCH GŁÓWNYCH I ANALIZA CZYNNIKOWA

Wprowadzenie

Analiza składowych głównych i analiza czynnikowa to dwie najbardziej popularne metody pozwalające na sprowadzenie dużej liczby badanych zmiennych do znacznie mniejszej liczby wzajemnie niezależnych składowych głównych lub czynników. Nowe zmienne (składowe główne lub czynniki) zachowują stosunkowo dużą część informacji zawartych w zmiennych pierwotnych, a jednocześnie każda z nich jest nośnikiem innych treści merytorycznych. Obie powyższe metody redukcji zmiennych są często stosowane, gdyż zbyt duża ilość rozpatrywanych cech powoduje wzrost skali trudności interpretacji.

Zasadniczą przyczyną podjęcia tematu jest próba pokazania, że wyżej wymienionych metod, choć są bardzo podobne, nie można utożsamiać. Mimo tego, iż w obu przypadkach są obliczane wartości własne, ładunki czynnikowe itp., to jednak istnieją między nimi różnice w sposobie działania, o czym należy pamiętać. Zatem stosowanie tych nazw zamiennie jest niedopuszczalne.

Artykuł składa się z trzech części. Rozdziały pierwszy i drugi są poświęcone, odpowiednio, analizie składowych głównych i analizie czynnikowej, gdzie została dokonana krótka charakterystyka tych metod. W rozdziale trzecim, na podstawie przykładu empirycznego, porównano efektywność analizy składowych głównych i analizy czynnikowej.

1. Analiza składowych głównych

Początki techniki analizy składowych głównych pochodzą od Pearsona (1901). Jednak główny rozwój tej metody zawdzięcza się pracom amerykańskiego statystyka Hotellinga (1933), który wykorzystał ją do analizy testów osiągnięć szkolnych.

Podstawową ideą metody jest transformacja wyjściowego zbioru zmiennych $X_1, \dots, X_p$ na nowy zbiór zmiennych $Z_1, \dots, Z_p$, zwanych składowymi głównymi. W konsekwencji liczba głównych składowych jest równa liczbie zmiennych pierwotnych. W praktyce nie ma to jednak dużego znaczenia, gdyż liczbę składowych głównych ogranicza się w dalszych rozważaniach do kilku najważniejszych. Zatem celem analizy składowych głównych jest redukcja liczby zmiennych przy zachowaniu tak dużej zmienności danych, jak to tylko możliwe.

Model matematyczny w analizie składowych głównych jest sformułowany w postaci następującego układu równań liniowych:

$$\begin{aligned} X_1 &= a_{11}Z_1 + a_{12}Z_2 + \cdots + a_{1p}Z_p \\ X_2 &= a_{21}Z_1 + a_{22}Z_2 + \cdots + a_{2p}Z_p \\ &\vdots \\ X_p &= a_{p1}Z_1 + a_{p2}Z_2 + \cdots + a_{pp}Z_p \end{aligned}$$

Zmienne rzeczywiste podlegające obserwacji X_i dla $i \in \{1, \dots, p\}$ są wyrażone jako kombinacje liniowe zmiennych nieobserwowalnych Z_j dla $j \in \{1, \dots, p\}$, zwanych składowymi głównymi. Współczynniki a_{ij} dla $i, j \in \{1, \dots, p\}$ określają wagę danej składowej w opisie zmiennych empirycznych.

1.1. Algorytm postępowania w analizie składowych głównych

Poniższe kroki opisują schemat postępowania w analizie składowych głównych [2; 5; 8].

Krok I – Sprawdzenie założeń

Przed rozpoczęciem analizy składowych głównych należy sprawdzić podstawowe założenie, aby ocenić zasadność jej zastosowania, a mianowicie skorelowanie zmiennych – im wyższe korelacje między zmiennymi pierwotnymi, tym bardziej uzasadnione jest wykorzystanie tej analizy. Korelację bada się analizując macierz korelacji dla zmiennych wziętych do analizy lub wykorzystując test Bartletta [8].

Należy również zwrócić uwagę na poniższe warunki [3; 8]:

1. **Normalność rozkładu** – założenie to nie jest konieczne, gdy analizuje się duży zbiór danych.
2. **Liczebność i reprezentatywność próby** – do analizy przystępuje się, gdy próba liczy co najmniej 50 obserwacji. Próbę należy pobrać w sposób losowy. Zbiór obserwacji musi być jednorodny.

3. **Punkty odstające** – punkty odstające niestety często zniekształcają prawdziwe zależności między zmiennymi. Dobrze jest na początku analizy wykryć takie punkty i usunąć je z danych.
4. **Braki danych** – w przypadku brakujących danych w analizowanej próbie należy zastąpić braki przez średnie lub usunąć przypadki z brakującymi danymi.

Krok II – Wybór odpowiedniej macierzy

Następnie należy przyjrzeć się początkowym zmiennym. Jeżeli analizowane zmienne są porównywalne (wyrażają się w tych samych jednostkach i są tego samego rzędu), to w dalszej analizie wykorzystuje się macierz kowariancji. Jeżeli natomiast zmienne mają różne jednostki lub są różnego rzędu, analizę składowych głównych przeprowadza się wykorzystując macierz korelacji. Jest to ważny krok rozpoczynający całą analizę, gdyż składowe główne otrzymane dla macierzy kowariancji i korelacji nie muszą być takie same.

Krok III – Wyznaczenie składowych głównych

Niech $\boldsymbol{X} = \left(X_1, \dots, X_p\right)^T$ będzie wektorem zmiennych wziętych do analizy. Składowe główne są kombinacją liniową zmiennych początkowych:

$$\begin{aligned} Z_1 &= a_{11}X_1 + a_{21}X_2 + \cdots + a_{p1}X_p \\ Z_2 &= a_{12}X_1 + a_{22}X_2 + \cdots + a_{p2}X_p \\ &\vdots \\ Z_p &= a_{1p}X_1 + a_{2p}X_2 + \cdots + a_{pp}X_p \end{aligned}$$

Kolejnym krokiem jest wyznaczenie macierzy współczynników a_{ij} dla $i, j \in \{1, \dots, p\}$ dla zadanego z góry wektora obserwacji $\boldsymbol{X}$. Algorytm wyznaczania współczynników a_{ij} dla $i, j \in \{1, \dots, p\}$ bardzo dokładnie opisuje D.F. Morrison [5] i A. Stanisz [8].

Krok IV – Redukcja wymiaru – kryteria wyboru

Ważną informacją jest to, że każda kolejna wyznaczona składowa główna wyjaśnia coraz mniejszą część zmienności początkowych zmiennych. W jakimś momencie okaże się, że któraś z kolei składowa określa znikomą część zmienności. Należy zatem dokonać redukcji składowych, stosując w dalszych rozważaniach tylko najważniejsze.

Popularne kryteria redukcji [2; 4; 8]:

1. **Kryterium wystarczającej proporcji** – stopień wyjaśnionej wariancji oryginalnych zmiennych musi wynosić co najmniej 75%. W praktyce najczęściej już przy 2-3 głównych składowych stopień wyjaśnienia wariancji jest wystarczający.

2. **Kryterium Kaisera** – eliminacja składowych głównych, których wartości własne są mniejsze od 1.
3. **Wykres osypiska** – wyznaczenie na wykresie liniowym kolejnych wartości własnych. Interpretacja polega na znalezieniu miejsca, od którego na prawo występuje łagodny spadek wartości własnych. Nie powinno się uwzględniać więcej czynników, niż te znajdujące się po lewej stronie tego punktu.

Wybór odpowiedniego kryterium leży w gestii statystyka, dlatego też decyzja ta jest dosyć subiektywna i wpływa na rezultaty analizy.

Krok V – Interpretacja

Interpretację otrzymanych wyników przeprowadza się za pomocą tzw. ładunków czynnikowych. Ładunki czynnikowe są współczynnikami korelacji pomiędzy daną zmienną a składowymi.

Jeżeli powyższa analiza jest przeprowadzana na podstawie macierzy kowariancji, to współczynnik korelacji pomiędzy i-tą zmienną X_i a j-tą składową Z_j dla $i, j \in \{1, \dots, p\}$ oblicza się ze wzoru:

$$r_{X_i,Z_j} = \frac{cov(X_i, Z_j)}{s_i\sqrt{\lambda_j}} = \frac{\lambda_j a_{ij}}{s_i\sqrt{\lambda_j}} = \frac{\sqrt{\lambda_j}a_{ij}}{s_i}$$

gdzie:
s_i – odchylenie standardowe zmiennej X_i,
λ_j – wariancja składowej głównej Z_j, a także j-ta co do wielkości wartość własna macierzy korelacji (kowariancji), na której opiera się cała analiza,
$\sqrt{\lambda_j}$ – odchylenie standardowe składowej Z_j.

Jeśli natomiast składowe są generowane z macierzy korelacji, to:

$$r_{X_i,Z_j} = \sqrt{\lambda_j}a_{ij}$$

Suma wszystkich wartości własnych macierzy korelacji (kowariancji) $\lambda_1 + \cdots + \lambda_p$ jest całkowitą wariancją układu. Dzięki temu można zdefiniować część całkowitej wariancji wyznaczoną przez j-tą składową:

$$h_j = \frac{\lambda_j}{\lambda_1 + \cdots + \lambda_p} \cdot 100\%$$

Natomiast procentowy udział zmienności całkowitej wyjaśnionej przez k pierwszych składowych oblicza się następująco:

$$H_k = \sum_{j=1}^{p} h_j$$

2. Analiza czynnikowa

Twórcami głównej koncepcji tej metody są psychologowie CH. Spearman (1904) i L.L. Thurstone (1913). Ch. Spearman wprowadził pojęcie pojedynczego czynnika ogólnego dla wyjaśnienia wyników testów inteligencji. Dopiero L.L. Thurstone stworzył podstawy teoretyczne analizy czynnikowej. Celem analizy czynnikowej jest dążenie do wyodrębnienia wszystkich czynników, które mogą rzeczywiście tkwić w korelacjach danego układu zmiennych, jednocześnie zachowując jak najwięcej informacji zawartych w zmiennych pierwotnych, a następnie redukcja tych czynników.

Model analizy czynnikowej konstruuje się jako założenie wstępne, które jest sformułowane w postaci układu równań:

$$\begin{aligned} X_1 &= a_{11}F_1 + a_{12}F_2 + \cdots + a_{1m}F_m + b_1U_1 \\ X_2 &= a_{21}F_1 + a_{22}F_2 + \cdots + a_{2m}F_m + b_2U_2 \\ &\vdots \\ X_p &= a_{p1}F_1 + a_{p2}F_2 + \cdots + a_{pm}F_m + b_pU_p \end{aligned}$$

gdzie $m < p$.

Standaryzowane zmienne pierwotne X_i dla $i \in \{1, \dots, p\}$ są wyrażone jako liniowe funkcje zmiennych nieobserwowalnych F_j dla $j \in \{1, \dots, m\}$, zwanych czynnikami wspólnymi i pojedynczego czynnika losowego U_i dla $i \in \{1, \dots, p\}$, zwanego czynnikiem swoistym. Współczynniki a_{ij} oraz b_i dla $i \in \{1, \dots, p\}$, $j \in \{1, \dots, m\}$ są zwane ładunkami czynnikowymi i określają wagę danego czynnika w opisie zmiennych empirycznych.

W analizie czynnikowej przyjmuje się dwa założenia o zmiennych i czynnikach:

1. Zmienne i czynniki są zestandaryzowane.
2. Czynniki wspólne F_j są ze sobą nieskorelowane, czynniki swoiste U_i są ze sobą nieskorelowane, czynniki wspólne F_j są nieskorelowane z czynnikami swoistymi U_i dla $i \in \{1, \dots, p\}$, $j \in \{1, \dots, m\}$.

2.1. Algorytm postępowania w analizie czynnikowej

Poniższe kroki opisują schemat postępowania w analizie czynnikowej [2; 5; 8; 9].

Krok I – Sprawdzenie założeń

Założenia w analizie czynnikowej są podobne jak w analizie składowych głównych z tym wyjątkiem, że zmienne pierwotne powinny mieć rozkład normalny lub być doprowadzone do takiej postaci drogą odpowiednich transformacji. Punktem wyjścia obliczeń jest macierz korelacji. Należy dokonać wstępnej oceny istniejących korelacji.

Krok II – Metody estymacji modelu analizy czynnikowej

Rozwiązanie analizy czynnikowej polega na wyznaczeniu układu czynników wspólnych F_j dla $j \in \{1, \dots, m\}$, co jest równoważne z określeniem dla każdego czynnika F_j odpowiadającego mu wektora $(a_{1j}, \dots, a_{pj})$. Dokonuje się tego wykorzystując jedną z podstawowych metod estymacji, do których m.in. należą [1; 5; 7; 8; 9]:

1. **Metoda głównych składowych** – opracowana przez Hotellinga (1933).
2. **Metoda głównego czynnika** – opracowana przez Harmana (1960).
3. **Metoda największej wiarygodności** – opracowana przez Lawleya (1940).
4. **Metoda centroidalna** – opracowana przez Thurstone'a (1931).

Największe uznanie matematyków zdobyła metoda głównych składowych. Nie bez przyczyny jest ona ustawiona jako metoda domyślna w programie Statistica w analizie czynnikowej. Wybór każdej z tych metod jest zawsze obciążony mniejszą czy większą dozą arbitralności.

Krok III – Redukcja wymiaru – kryteria wyboru

Kryteria redukcji liczby czynników są analogiczne jak w analizie składowych głównych. Natomiast opierając analizę czynnikową na metodzie największej wiarygodności, można za pomocą istniejącego testu dobroci dopasowania określić, czy ilość wybranych czynników jest właściwą liczbą dla danego modelu, czy też nie [2; 5; 9].

Krok IV – Rotacja czynników

Często zdarza się, że zmienna ma wysokie ładunki na kilku czynnikach, co uniemożliwia jednoznaczną interpretację. W takiej sytuacji należy przeprowadzić rotację czynników. W większości przypadków rotacja czynników redukuje dwuznaczność interpretacji, jaka może wystąpić w rozwiązaniu bez rotacji.

Dzięki obrotowi można łatwiej utożsamić każdy czynnik ze zmiennymi, z którymi jest mocno skorelowany.

Ustalenie najwłaściwszej pozycji układu odniesienia jest jednym z najtrudniejszych kroków. Według L.L. Thurstone'a (1935) należy dążyć do tzw. **prostej struktury**, która znacznie ułatwia interpretację wyników. Prostota takiej struktury ładunków czynnikowych polega na tym, że każda zmienna ma stosunkowo najprostszą zawartość czynnikową, tj. dominujący ładunek jakiegoś jednego czynnika i odwrotnie – miarą danego czynnika są tylko niektóre spośród analizowanych zmiennych. W praktyce rzadko można doprowadzić do struktury czynnikowej spełniającej kryteria struktury prostej, należy jednak dążyć do uzyskania wyniku najbardziej do niej zbliżonego.

Do wykonania rotacji najczęściej stosuje się metodę VARIMAX lub QUARTIMAX [5; 9], które ostatecznie decydują o interpretacji modelu, gdyż różne metody dają różne pozycje układów osi czynników.

Krok V – Interpretacja

Podstawowym zadaniem analizy czynnikowej jest wyznaczenie macierzy współczynników a_{ij} zwanych ładunkami czynnikowymi dla $i \in \{1, \dots, p\}$, $j \in \{1, \dots, m\}$. Ładunki te można interpretować w ten sposób, że waga czynnika jest współczynnikiem korelacji między zmienną a czynnikiem. Zatem:

$$r_{X_i F_j} = a_{ij}$$

$$\text{dla } i \in \{1, \dots, p\}, j \in \{1, \dots, m\}.$$

Do interpretacji otrzymanych wyników szuka się tych zmiennych, które mają najwyższe (w wartościach bezwzględnych) wartości ładunków czynnikowych dla danych czynników. Ładunki czynnikowe opisują wkład zmiennej do poszczególnych czynników.

Część całkowitej wariancji wyjaśnionej przez j-ty czynnik jest obliczany ze wzoru:

$$h_j = \frac{\lambda_j}{\lambda_1 + \dots + \lambda_m} \cdot 100\%$$

gdzie:
λ_j – j-ta wartość własna macierzy korelacji dla $j \in \{1, \dots, m\}$.

Natomiast procentowy udział zmienności całkowitej wyjaśnionej przez k pierwszych czynników oblicza się następująco:

$$H_k = \sum_{j=1}^{m} h_j$$

3. Porównanie efektywności analizy składowych głównych i analizy czynnikowej

3.1. Informacje wstępne

Do badań posłużyły dane z Rocznika Statystycznego Pracy 2010. Analizie poddano 311 powiatów Polski ze względu na osiem zmiennych:

1. Bezrobotni poprzednio pracujący – BPP.
2. Bezrobotni zwolnieni z przyczyn dotyczących zakładów pracy – BZ.
3. Bezrobotni zamieszkali na wsi – BZW.
4. Bezrobotni nieposiadający prawa do zasiłku – BNPZ.
5. Zatrudnieni w warunkach zagrożenia związanego ze środowiskiem pracy – ZŚP.
6. Zatrudnieni w warunkach zagrożenia związanego z uciążliwością pracy – ZUP.
7. Zatrudnieni w warunkach zagrożenia związanego z czynnikami mechanicznymi – ZCM.
8. Poszkodowani w wypadku przy pracy – PPP.

3.2. Wyniki analizy empirycznej

W artykule tym dokonano redukcji liczby zmiennych opisujących zróżnicowanie powiatów Polski. Uzyskane wyniki pozwalają na porównanie metody analizy składowych głównych oraz analizy czynnikowej, wskazując przede wszystkim stopień efektywności każdej z nich. Obie analizy zostały przeprowadzone za pomocą programu Statistica.

Przeprowadzając badanie za pomocą analizy składowych głównych, wykorzystano macierz korelacji i otrzymano następujące rezultaty:

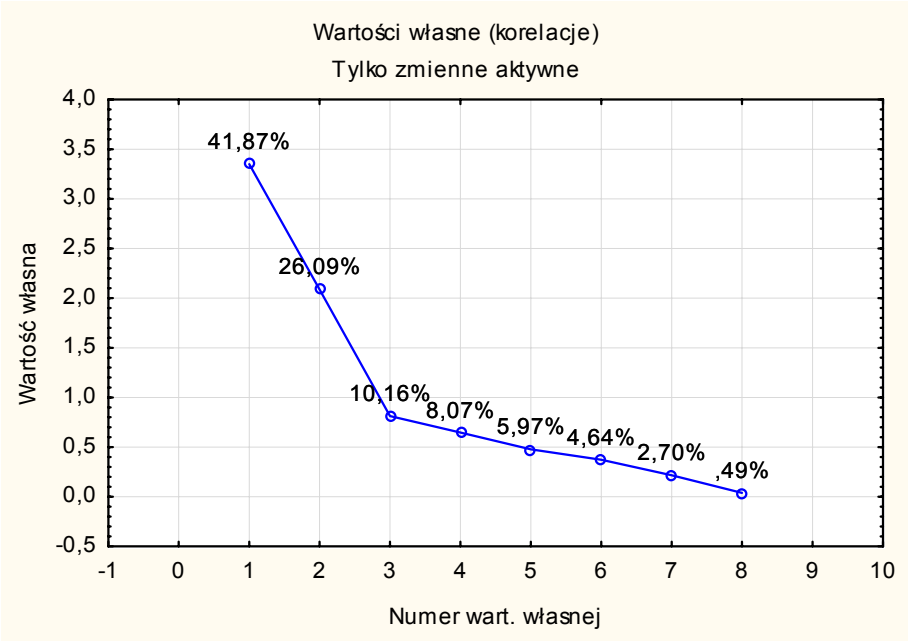
1. Korzystając z kryterium wystarczającej proporcji, dokonano redukcji ośmiu zmiennych do trzech składowych głównych. Na podstawie tabeli 1 można odczytać, iż pierwsza składowa wyjaśnia niecałe 42% całkowitej zmienności. Druga składowa wyjaśnia 26%, a trzecia 10%, co daje łącznie ponad 78% całkowitej zmienności.

Tabela 1

Wartości własne wyznaczone dla analizowanych danych, procent całkowitej wariancji wyjaśnionej przez j-tą składową (h_j), skumulowane wartości własne oraz skumulowany procent wyjaśnionej wariancji (H_k)

	Wartość własna	h_j (%)	Skumulowana wartość własna	H_k (%)
Z_1	3,349941	41,87426	3,34994	41,87426
Z_2	2,087453	26,09316	5,43739	67,96743
Z_3	0,813020	10,16275	6,25041	78,13018
Z_4	0,645555	8,06943	6,89597	86,19961
Z_5	0,477604	5,97005	7,37357	92,16966
Z_6	0,371222	4,64027	7,74479	96,80993
Z_7	0,215750	2,69689	7,96055	99,50681
Z_8	0,039455	0,49319	8	100

Wykres osypiska poniżej potwierdza tę decyzję (rysunek 1).



Rys. 1. Wykres osypiska

Źródło: Opracowanie własne z wykorzystaniem programu Statistica.

2. Ładunki czynnikowe traktuje się jako korelacje między zmiennymi a składowymi. Dla trzech pierwszych składowych ładunki przedstawiono w tabeli 2.

Tabela 2

Ładunki czynnikowe dla trzech pierwszych składowych

	Z_1	Z_2	Z_3
BPP	**-0,833739**	0,449623	0,073746
BZ	-0,511962	0,137246	**-0,818869**
BZW	-0,564090	**0,673689**	0,213194
BNPZ	**-0,826374**	0,505603	0,112882
ZŚP	-0,549488	**-0,655135**	-0,003492
ZUP	-0,534307	-0,460750	0,279763
ZCM	-0,539138	-0,511579	-0,020974
PPP	**-0,716612**	-0,503744	0,011093

Pierwsza składowa ma najwyższe, ujemnie ładunki czynnikowe ze zmiennymi BPP, BNPZ, PPP. Określa ona zatem bezrobotnych poprzednio pracujących, bezrobotnych nieposiadających prawa do zasiłku, poszkodowanych w wypadku przy pracy. Druga składowa ma najwyższe ładunki ze zmiennymi BZW i ZŚP, lecz korelacje te nie są zbyt wysokie. Obie zmienne oddziałują w sposób przeciwny na tą składową, BZW dodatnio, a ZŚP ujemnie. Trzecia składowa najsilniej i ujemnie jest związana ze zmienną BZ. Brakuje natomiast składowej najmocniej skorelowanej ze zmiennymi ZUP i ZCM. Obie te zmienne mają podobne (w wartościach bezwzględnych) wartości ładunków dla dwóch składowych – pierwszej i drugiej. Opisana struktura jest daleka od spełnienia warunków tzw. prostej struktury.

Warto sprawdzić, czy dodanie czwartej składowej głównej nie poprawi powyższej sytuacji. Jest to dość ryzykowne posunięcie, znacznie wpływające na rezultaty, gdyż jedynie kryterium wystarczającej proporcji jest spełnione. Wyniki przedstawia tabela 3.

Tabela 3

Ładunki czynnikowe dla czterech pierwszych składowych

	Z_1	Z_2	Z_3	Z_4
1	2	3	4	5
BPP	**-0,833739**	0,449623	0,073746	-0,012254
BZ	-0,511962	0,137246	**-0,818869**	0,179866
BZW	-0,564090	**0,673689**	0,213194	-0,086882
BNPZ	**-0,826374**	0,505603	0,112882	-0,013554

cd. tabeli 3

1	2	3	4	5
ZŚP	-0,549488	**-0,655135**	-0,003492	-0,028257
ZUP	-0,534307	-0,460750	0,279763	**0,561679**
ZCM	-0,539138	-0,511579	-0,020974	**-0,537121**
PPP	**-0,716612**	-0,503744	0,011093	-0,023245

Dodana czwarta składowa główna faktycznie jest w największym stopniu związana ze zmiennymi ZUP i ZCM, ale mimo wszystko w niezbyt wysokim. Krok ten jeszcze bardziej oddalił od spełnienia warunków prostej struktury, gdyż zmienne ZUP i ZCM mają teraz podobne (w wartościach bezwzględnych) wartości ładunków dla trzech składowych.

Przeprowadzając badanie za pomocą analizy czynnikowej, do wyodrębnienia czynników wykorzystano cztery metody: głównych składowych, głównego czynnika, największej wiarygodności oraz centroidalną. Za każdym razem wyniki są poprawione za pomocą rotacji Varimax. Wnioski przedstawiono poniżej:

1. Poniższe tabele zawierają: wartości własne wyznaczone dla analizowanych danych, procent całkowitej wariancji wyjaśnionej przez j-tą składową (h_j), skumulowane wartości własne oraz skumulowany procent wyjaśnionej wariancji (H_k) wyliczone za pomocą wymienionych wyżej metod.

Tabela 4

Metoda składowych głównych

	Wartość własna	h_j (%)	Skumulowana wartość własna	H_k (%)
F_1	3,349941	41,87426	3,34994	41,87426
F_2	2,087453	26,09316	5,43739	67,96743
F_3	0,813020	10,16275	6,25041	78,13018
F_4	0,645555	8,06943	6,89597	86,19961

Tabela 5

Metoda głównego czynnika

	Wartość własna	h_j (%)	Skumulowana wartość własna	H_k (%)
F_1	3,04155	38,01933	3,04155	38,01933
F_2	1,69275	21,15934	4,73429	59,17867
F_3	0,10162	1,27021	4,83591	60,44889
F_4	0,02170	0,27129	4,85761	60,72017

Tabela 6

Metoda największej wiarygodności

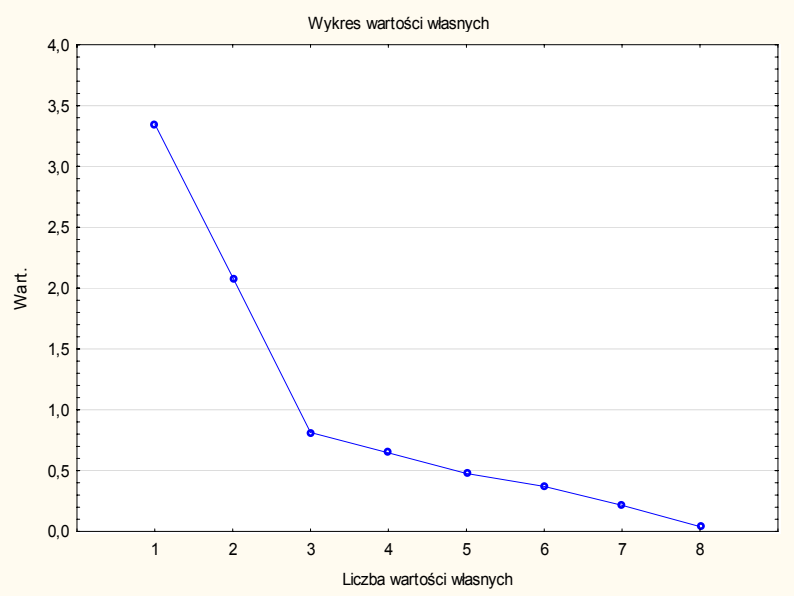
	Wartość własna	h_j (%)	Skumulowana wartość własna	H_k (%)
F_1	2,89747	36,21835	2,89747	36,21835
F_2	1,97842	24,73027	4,87589	60,94862
F_3	0,27838	3,47970	5,15427	64,42832
F_4	0,09061	1,13264	5,24488	65,56097

Tabela 7

Metoda centroidalna

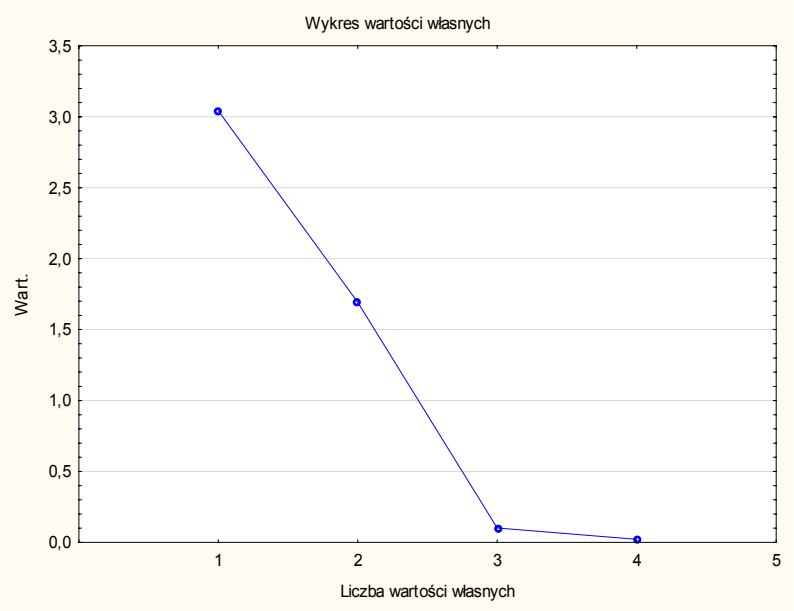
	Wartość własna	h_j (%)	Skumulowana wartość własna	H_k (%)
F_1	3,08785	38,59812	3,08785	38,59812
F_2	1,80870	22,60878	4,89655	61,20690
F_3	0,14521	1,81512	5,04176	63,02201
F_4	0,11341	1,41763	5,15517	64,43964

2. Wykresy osypiska:



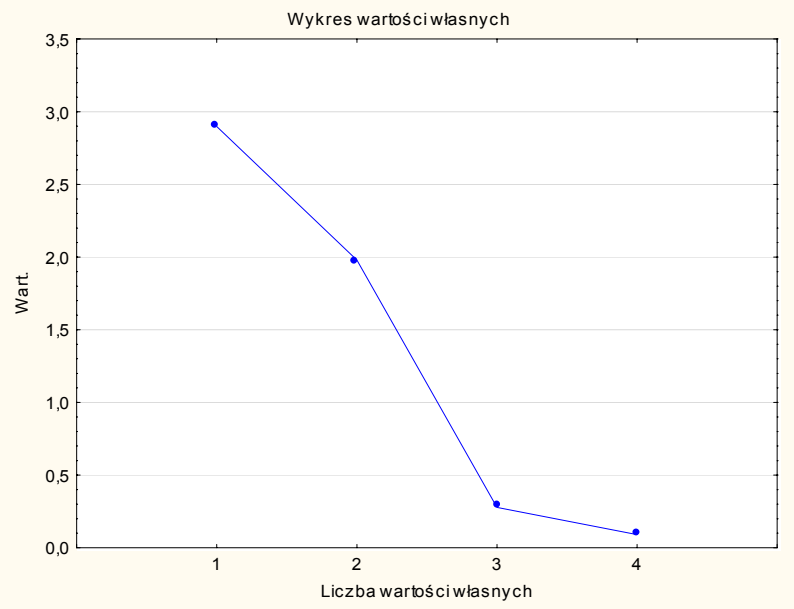
Rys. 2. Wykres osypiska – metoda składowych głównych

Źródło: Opracowanie własne z wykorzystaniem programu Statistica.



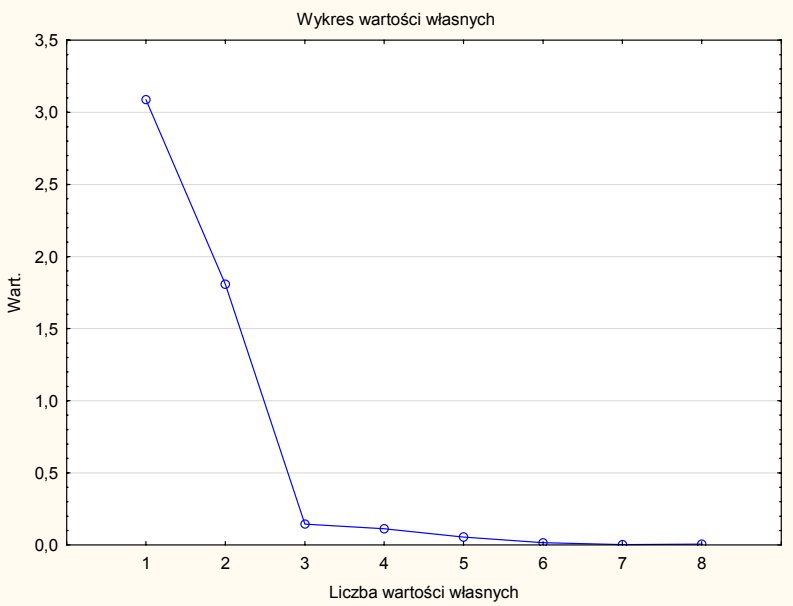
Rys. 3. Wykres osypiska – metoda głównego czynnika

Źródło: Opracowanie własne z wykorzystaniem programu Statistica.



Rys. 4. Wykres osypiska – metoda największej wiarygodności

Źródło: Opracowanie własne z wykorzystaniem programu Statistica.



Rys. 5. Wykres osypiska – metoda centroidalna

Źródło: Opracowanie własne z wykorzystaniem programu Statistica.

Zgodnie z wybranymi dwoma kryteriami – kryterium osypiska i kryterium wystarczającej proporcji, w którym dąży się do jak największego stopnia wyjaśnionej wariancji oryginalnych zmiennych – dokonano redukcji ośmiu zmiennych do trzech czynników w każdej z wybranych metod.

3. Tablice z wyodrębnionymi ładunkami czynnikowymi za pomocą wybranych metod z zastosowaną rotacją Varimax przedstawiono poniżej.

Tabela 8

Metoda składowych głównych

	F_1	F_2	F_3
BPP	**0,906088**	0,193881	0,210087
BZ	0,231644	0,105452	**0,941652**
BZW	**0,894938**	-0,128081	0,013673
BNPZ	**0,947446**	0,151940	0,174731
ZŚP	-0,028337	**0,846350**	0,118490
ZUP	0,162886	**0,727659**	-0,141502
ZCM	0,051718	**0,727890**	0,142561
PPP	0,192596	**0,838455**	0,165270

Tabela 9

Metoda głównego czynnika

	F_1	F_2	F_3
BPP	**0,937549**	0,182561	-0,150723
BZ	0,364813	0,191573	-0,131111
BZW	**0,785418**	-0,133661	0,233355
BNPZ	**0,971192**	0,138627	0,030187
ZŚP	0,028123	**0,783307**	0,021218
ZUP	0,127301	**0,571153**	0,044913
ZCM	0,103189	**0,618082**	0,023434
PPP	0,227712	**0,804953**	-0,095227

Tabela 10

Metoda największej wiarygodności

	F_1	F_2	F_3
BPP	**0,952829**	0,172299	-0,202474
BZ	0,367004	0,190854	-0,079508
BZW	**0,787298**	-0,134078	0,253331
BNPZ	**0,981287**	0,140274	0,064358
ZŚP	0,021719	**0,830838**	0,116879
ZUP	0,130949	**0,571697**	0,005766
ZCM	0,089348	**0,616609**	-0,033913
PPP	0,234874	**0,825477**	-0,181597

Tabela 11

Metoda centroidalna

	F_1	F_2	F_3
BPP	**0,932927**	0,186528	0,175063
BZ	0,361677	0,190384	0,147319
BZW	**0,785637**	-0,131191	-0,167420
BNPZ	**0,990931**	0,135145	-0,053094
ZŚP	0,025362	**0,811870**	-0,079666
ZUP	0,124214	**0,573591**	-0,115914
ZCM	0,106654	**0,616626**	0,006779
PPP	0,220190	**0,837739**	0,134804

Uzyskane wyniki wskazują, że najefektywniejszą metodą redukcji zmiennych w analizie czynnikowej jest metoda głównych składowych z zastosowaną

rotacją Varimax. Wybór trzech czynników w tej metodzie pozwolił na wyjaśnienie 78% całkowitej zmienności, co spełnia kryterium wystarczającej proporcji. Metoda ta w najlepszym stopniu przybliżyła wyniki analizy do tzw. prostej struktury, każda zmienna jest wysoko skorelowana tylko z jednym czynnikiem.

W pozostałych trzech metodach wybór trzeciego czynnika wydaje się zbędny. Niewiele on wnosi do wyjaśnienia całkowitej zmienności, która sięga mimo wszystko znacznie poniżej wymaganego poziomu 75%. Wyznaczone czynniki zachowują stosunkowo niedużą część informacji zawartych w zmiennych pierwotnych. Co więcej, zmienna BZ nie jest powiązana z żadnym czynnikiem, nawet dodanie czwartego czynnika nie zmieniłoby tej sytuacji.

Interpretując zatem wyniki analizy czynnikowej za pomocą metody składowych głównych, czynnik pierwszy wykazuje najwyższe ładunki dla zmiennych BPP, BZW oraz BNPZ, a więc jest związany głównie z bezrobociem. Czynnik drugi jest najwyżej skorelowany ze zmiennymi ZŚP, ZUP, ZCM oraz PPP, a więc jest związany z zatrudnieniem w warunkach zagrożenia i poszkodowaniem w wypadkach przy pracy, ogólnie dotyczy ciężkich warunków pracy. Czynnik trzeci, najsilniej związany ze zmienną BZ, również dotyczy bezrobotnych, ale konkretnie bezrobotnych zwolnionych. Można się zatem pokusić o następujące nazwy dla opisanych czynników:

– czynnik pierwszy – „Bezrobocie”,
– czynnik drugi – „Ciężkie warunki pracy”,
– czynnik trzeci – „Zwolnienie”.

Podsumowanie

Celem artykułu było porównanie efektywności analizy składowych głównych i analizy czynnikowej. Obie metody służą do redukcji zmiennych oraz do wyjaśniania istniejących korelacji między zmiennymi za pomocą kilku nieobserwowalnych i nieskorelowanych składowych głównych czy czynników. Do badań posłużyły dane z Rocznika Statystycznego Pracy 2010. Analizie poddano 311 powiatów Polski ze względu na osiem zmiennych, które w konsekwencji w analizie składowych głównych oraz analizie czynnikowej zostały zredukowane do trzech składowych i trzech czynników.

Analiza wybranego przykładu wykazała, iż wyniki otrzymane drogą analizy czynnikowej wykorzystującej metodę głównych składowych łatwiej poddają się interpretacji niż wyniki analizy składowych głównych. Wpływ na to ma niewątpliwie możliwość wykorzystania rotacji. W tym przypadku analiza czynnikowa okazała się efektywniejsza.

Literatura

1. Czyż T.: *Zastosowanie metody analizy czynnikowej do badania ekonomicznej struktury regionalnej Polski*. Wydawnictwo Polskiej Akademii Nauk, Wrocław 1971.
2. Frątczak E.: *Wielowymiarowa analiza statystyczna. Teoria – przykłady zastosowań z systemem SAS*. Szkoła Główna Handlowa, Warszawa 2009.
3. Grabiński T.: *Metody taksonometrii*. Akademia Ekonomiczna, Kraków 1992.
4. Krzyśko M.: *Wielowymiarowa analiza statystyczna*. Wydawnictwo Naukowe UAM, Poznań 2000.
5. Morrison D.F.: *Wielowymiarowa analiza statystyczna*. Państwowe Wydawnictwo Naukowe, Warszawa 1990.
6. Pluta W.: *Wielowymiarowa analiza porównawcza w modelowaniu ekonometrycznym*. Państwowe Wydawnictwo Naukowe, Warszawa 1986.
7. Pluta W.: *Wielowymiarowa analiza porównawcza w badaniach ekonomicznych*. Państwowe Wydawnictwo Ekonomiczne, Warszawa 1977.
8. Stanisz A.: *Przystępny kurs statystyki z zastosowaniem Statistica PL na przykładach z medycyny*. T. 3: *Analizy wielowymiarowe*. StatSoft, Kraków 2007.
9. Walesiak M., Gatnar E.: *Statystyczna analiza danych z wykorzystaniem programu R*. Wydawnictwo Naukowe PWN, Warszawa 2009.

COMPARATIVE ANALYSIS OF EFFECTIVENESS OF THE METHODS FOR REDUCTION OF VARIABLES – PRINCIPAL COMPONENT ANALYSIS AND FACTOR ANALYSIS

Summary

Principal component analysis and factor analysis are the two most popular methods that allow to bring a large number of studied variables to a much smaller number of mutually independent principal components or factors. New variables (principal components or factors) retain a relatively large part of the information contained in the original variables, while each of them is a carrier of other substantive content. Both of these methods of reduction of the variables are often used, because too many pending attributes increases the range of the difficulty of interpretation.

The main reason of undertaking the project is an attempt to show, that the abovementioned methods, although they are very similar, cannot be indentified. Despite the fact, that in both cases eigenvalues are calculated, factor loadings, etc., but still there are differences in the way of action, about which it must be remembered. So the usage of these names the variables are unacceptable.

The article consists of three parts. The first and second chapter are devoted, respectively, to the analysis of the principal components and factor analysis, where a short characterization of these methods had been made. In the third chapter, on the basis of an empirical example, we compared the effectiveness of the principal components analysis and factor analysis.

Katarzyna Jakowska-Suwalska

WIELOKRYTERIALNY, NIELINIOWY MODEL WIELKOŚCI ZAMÓWIENIA MATERIAŁÓW DLA KOPALNI WĘGLA KAMIENNEGO*

Wprowadzenie

W teorii sterowania zapasami występuje wiele modeli, które pozwalają ustalić politykę ustalania zapasów i wyznaczania wielkości zamówienia. W większości modeli jako kryterium oceny rozwiązań używa się funkcji kosztów (zamawiania i utrzymania zapasów) [11; 8].

W pracach [3; 4] przedstawiono wielokryterialne modele, na podstawie których można wyznaczyć wielkości:

- zamówienia,
- terminu zamówienia,
- zapasów magazynowych,

gdzie jako funkcji skalaryzującej użyto funkcji kosztów związanych z wielkością zamówienia, zapasów magazynowych oraz brakiem materiału do produkcji. W kopalniach węgla kamiennego wchodzących w skład Kompanii Węglowej S.A. wielkość zamówienia podlegającego ustawie o zamówieniach publicznych planuje się około roku wcześniej. Jest to związane z czasem ustalenia planów zakupów dla wszystkich kopalni oraz z czasem postępowania przetargowego. Zatem wielkość zamówienia materiału dla kopalni należy wyznaczyć jednorazowo na podstawie planów finansowych oraz planów wydobycia na następny rok. Do rozwiązania tego problemu zaproponowano wielokryterialny model wielkości zamówienia dla materiałów, których zużycie, a więc także zapotrzebowanie jest zmienną losową o znanym rozkładzie prawdopodobieństwa.

* Praca powstała w ramach realizacji projektu badawczego nr N N524 552038 „Wielokryterialne wspomaganie planowania i kontrolowania potrzeb materiałowych w przedsiębiorstwie górniczym" finansowanego przez Ministerstwo Nauki i Szkolnictwa Wyższego.

1. Konstrukcja wielokryterialnego modelu wielkości zamówienia

Niech X_i będzie zmienną losową o znanej dystrybuancie F_i, oznaczającą wielkość zużycia materiału M_i na tonę wydobycia, natomiast z_i poszukiwaną wielkością zamówienia materiału M_i na tonę wydobycia (i = 1, 2,…, s). Zgodnie z teorią zapasów należy zamówić taką ilość z_i materiału M_i, aby z jak największym prawdopodobieństwem pokryła ona popyt na ten materiał. Wiadomo, że zamrożony w magazynie materiał zwiększa koszty przedsiębiorstwa. Należy więc zamawiać taką ilość materiału, aby wielkość zakupu była jak najmniejsza, lecz nie odchylała się zbytnio od przeszłych wielkości zapotrzebowania na materiał, natomiast koszty zakupu wszystkich materiałów nie przekraczały pewnej zadanej kwoty K. Założono, że wielkość zużycia na materiał M_i (i = 1, 2,…, s) nie wykazuje trendu ani wahań okresowych. Przyjęto także, że wszystkie wiadomości o warunkach panujących w kopalni, mających wpływ na wielkości zużycia materiałów, znajdują się w danych z przeszłych okresów.

Za funkcje kryteria przyjęto dla każdego materiału M_i:

- wielkość zamówienia z_i,
- wielkości odchyleń wielkości zamówienia z_i od rzeczywistych wielkości zużycia materiału x_{i1}, x_{i2},… ,x_{in} w ostatnich n okresach,
- prawdopodobieństwo braku materiału M_i do wykonania robót.

Model ten można zapisać w postaci:

$$\begin{cases} z_i \to \min, \\ F_i(z_i) \to \max, \\ |x_{it} - z_i| \to \min, \\ \sum_{i=1}^{s} c_i z_i \le K, \\ z_i \ge 0, \\ i = 1, 2, ..., s, \\ t = 1, 2, ..., n, \end{cases} \tag{1}$$

gdzie:

c_i – cena jednostki materiału M_i,

K – kwota przeznaczona na zakup materiałów M_1, M_2,…, M_s.

Przyjęto oznaczenie:

$$f_{celu}(z_1, z_2, \ldots, z_s) = (F_1(z_1), F_2(z_2), \ldots, F_s(z_s), z_1, z_2, \ldots, z_s, |x_{1t} - z_1|, \ldots, |x_{st} - z_s|, t = 1, 2, ., n).$$

W celu wyznaczenia rozwiązań efektywnych wielokryterialnego problemu najczęściej wprowadza się skalaryzację zagadnienia [1; 5; 6; 7]. W przypadku rozważanego modelu będzie ona miała postać:

$$\max(s(u, f_{celu}(z_{1,} z_2,...,z_s) : z_{1,} z_2,...,z_s \in Q) \quad u \in U,$$

gdzie:
u – wektor parametrów sterujących,
$s : U \times Y \to R$ – funkcja skalaryzująca,
Q – zbiór ograniczeń.

Jeśli oznaczy się przez u_i wagę nadaną przez decydenta materiałowi M_i ($i = 1, 2,..., s$) na podstawie ważności materiału w procesie produkcyjnym, tak aby $u_1, u_2, ...u_s > 0$ oraz $u_1 + u_2 + ... + u_s = 1$, to można przeprowadzić skalaryzację za pomocą średnich ważonych. Model wtedy przyjmie postać:

$$\begin{cases} \sum_{i=1}^{s} u_i z_i \to \min & (a) \\ \sum_{i=1}^{s} u_i F_i(z_i) \to \max & (b) \\ \sum_{i=1}^{s} u_i |x_{it} - z_i| \to \min & (c) \\ t = 1,2,...,n \\ \sum_{i=1}^{s} c_i z_i \le K \\ z_i \ge 0 \\ i = 1,2,...,s. \end{cases}$$

Ponieważ w grupie kryteriów (c) występują różnice $|x_{it} - z_i|$ pomiędzy wielkościami zamówienia materiału z_i a jego zużyciem x_{it} w minionych okresach $t = 1,2,...,n$, więc warto przeprowadzić proces postarzania obserwacji poprzez wprowadzenie dla poszczególnych okresów odpowiednich wag. Można zastosować jeden z następujących sposobów wyznaczania wag:

– $w_t = \dfrac{2t}{n(n+1)}$ dla $t = 1, 2,..., n$ (wagi liniowe),

– $w_t = w_{t-1} + \dfrac{1}{n(n+1-t)}$ dla $t = 1, 2,..., n$ przy $w_0 = 0$ (wagi harmoniczne),

– $w_{n-t+1} = \dfrac{(1-a)^t}{\sum_{t=1}^{n}(1-a)^t}$ dla $t = 1, 2,..., n$ oraz $a \in [0,1]$ (wagi wykładnicze).

Dla grupy kryteriów (c) można więc wprowadzić skalaryzację następująco:

$$\sum_{t=1}^{n}\sum_{i=1}^{s} u_i w_t \left| x_{it} - z_i \right|.$$

Jeśli decydent wyznaczy wagi $\alpha_{(a)}, \alpha_{(b)}, \alpha_{(c)} \geq 0$ ustalające ważność poszczególnych grup kryteriów (*a*), (*b*), (*c*), można wprowadzić skalaryzację w postaci:

$$s(u, f_{celu}(z_1, z_2, ..., z_s)) = \frac{\alpha_{(b)} \sum_{i=1}^{s} u_i F_i(z^u{}_i)}{\alpha_{(a)} \sum_{i=1}^{s} u_i z^u{}_i + \alpha_{(c)} \sum_{t=1}^{n} \sum_{i=1}^{s} u_i w_t \left(\left| x_{it}{}^u - z^u{}_i \right| \right)}, \quad (2)$$

gdzie:

$z^u{}_i$ – zunitaryzowane wielkości zamówienia z_i,

$x_{it}{}^u$ – zunitaryzowane wartości wielkości zużycia.

Do unitaryzacji zostanie wykorzystana reguła [7]:

$$y_t^u = \frac{y - \min(y_t : t = 1, 2, ..., n)}{\max(y_t : t = 1, 2, ..., n) - \min(y_t : t = 1, 2, ..., n)}. \quad (3)$$

Wartości $z^u{}_i$, $x^u{}_{it}$ dla $i = 1, 2, ..., s$ wyznaczono ze wzorów:

$$z^u{}_i = \frac{z_i}{\max(x_{i_t} : t = 1, 2, ..., n) - \min(x_{i_t} : t = 1, 2, ..., n)}, \quad i = 1, 2, ...s,$$

$$x_{it}^u = \frac{x_{it}}{\max(x_{it} : t = 1, 2, ..., n) - \min(x_{it} : t = 1, 2, ..., n)}, \qquad i = 1, 2, ...s$$

gdzie x_{it} to obserwacje wielkości zużycia materiału M_i w okresach $t = 1, 2, \ldots n$. Model (1) przyjmie więc postać:

$$\begin{cases} \dfrac{\alpha_{(b)} \sum_{i=1}^{s} u_i F_i(z^u{}_i)}{\alpha_{(a)} \sum_{i=1}^{s} u_i z^u{}_i + \alpha_{(c)} \sum_{t=1}^{n} \sum_{i=1}^{s} u_i w_t \left(\left| x^u{}_{it} - z^u{}_i \right| \right)} \to \max \\ \max(x_{it} : t = 1,2,...,n) \geq z_i \geq 0 \\ i = 1,2,...,s \\ \sum_{i=1}^{s} c_i z_i \leq K \\ u_1, u_2, ...u_s > 0, \; u_1 + u_2 + ... + u_s = 1 \\ \alpha_{(a)}, \alpha_{(b)}, \alpha_{(c)} > 0, \;\; \alpha_{(a)} + \alpha_{(b)} + \alpha_{(c)} = 1 \end{cases} \tag{4}$$

2. Wskaźniki oceny rozwiązań

W artykule przyjęto, że wszystkie informacje o procesie produkcyjnym, warunkach panujących w kopalni mających wpływ na zużycie materiałów są zawarte w danych historycznych dotyczących wielkości wydobycia oraz wielkości zużycia materiałów. Założono, że w planowanym okresie warunki te się nie zmienią. Do oceny wielkości zamówienia z_i materiału M_i zaproponowano wskaźniki oparte na danych historycznych:

– wskaźnik nadmiaru i braku materiału M_i (W_{nbi}),
– wskaźnik sumy nadmiaru i braku materiału M_i (W_{sbi}).

Zostały one zdefiniowane na podstawie prostych wskaźników zysków i strat [2] stosowanych na rynkach finansowych:

$$W_{nbi} = \frac{count\ G_i}{count\ L_i}, \tag{5}$$

$$W_{sbi} = \frac{\sum_{t=1}^{n} G_{it}}{\sum_{t=1}^{n} L_{it}}, \tag{6}$$

gdzie:

$$G_{it} = \begin{cases} \left| x_{it} - z_i \right| & gdy\ x_{it} - z_i \leq 0 \\ 0 & gdy\ x_{it} - z_i > 0, \end{cases} \tag{7}$$

$$L_{it} = \begin{cases} x_{it} - z_i & gdy \quad x_{it} - z_i > 0 \\ 0 & gdy \quad x_{it} - z_i \le 0, \end{cases} \tag{8}$$

x_{it} – obserwacje wielkości zużycia materiału M_i w okresach $t = 1, 2,\ldots n$,
count G_i – liczba różnych od zera wartości G_{it},
count L_i – liczba różnych od zera wartości L_{it}.

Jeśli wskaźnik $W_{nbi} \ge 1$, to oznacza, że przy wielkości zamówienia z_i w minionych okresach częściej występowałby nadmiar materiału M_i niż jego brak. Zakładając niezmienność warunków, można wtedy stwierdzić, że w planowanym okresie ta tendencja się utrzyma.

Jeśli wskaźnik $W_{sbi} \ge 1$, to oznacza, że przy wielkości zamówienia z_i w minionych okresach sumaryczny nadmiar materiału M_i (zapasy) byłby większy niż jego sumaryczny brak. Wynika stąd, że jeśli wskaźnik W_{sbi} ma wysoką wartość (większą niż 1), to przy zamówieniu z_i narastałyby zapasy materiału M_i.

3. Przykład zastosowania wielokryterialnego modelu dla ustalenia wielkości zamówień na drewno kopalniane i klej poliuretanowy

Drewno kopalniane jest zużywane w trakcie robót eksploatacyjnych do tworzenia obudów wyrobisk kopalnianych, natomiast klej poliuretanowy do uszczelniania wyrobisk [10]. Na podstawie medianowego testu serii na poziomie istotności 0,05 stwierdzono, że nie ma podstaw do odrzucenia hipotezy o braku trendu miesięcznych wielkości zużycia drewna kopalnianego w m^3 na tonę wydobycia oraz kleju poliretanowego w kg na tonę wydobycia. Wykazano także, że wielkości zużycia obu tych materiałów nie wykazują wahań okresowych. Można więc przyjąć, że miesięczne zużycie drewna oraz kleju poliuretanowego wykazują stały średni poziom z wahaniami przypadkowymi. W tabelach 1 i 2 podano podstawowe parametry rozkładu zużycia drewna i kleju poliuretanowego wyznaczone na podstawie miesięcznych danych z lat 2008-2010.

Tabela 1

Podstawowe parametry rozkładu miesięcznego zużycia drewna

Parametry	Zużycie drewna w m^3/t
1	2
Wartość maksymalna	0,006892
Wartość minimalna	0,002427
Wartość średnia *m*	0,003580

cd. tabeli 1

1	2
Odchylenie standardowe σ	0,00086
Współczynnik zmienności	0,238
Mediana	0,003506
Kurtoza	4,796
Skośność	1,687

Źródło: Opracowanie własne na postawie danych z kopalni zrzeszonej w Kompanii Węglowej S.A.

Wielkości zużycia drewna wykazują silną koncentrację wokół średniej oraz lewostronną asymetrię. Na podstawie wielkości zużycia drewna (w m^3/t) w ostatnich trzech latach stwierdzono (testem Kołmogorowa-Smirnowa na poziomie istotności 0,05), że jest ono zmienną losową o rozkładzie normalnym $N(0{,}00358;0{,}0086)$.

Tabela 2

Podstawowe parametry rozkładu miesięcznego zużycia kleju poliuretanowego

Parametry	Zużycie kleju poliuretanowego w kg/t
Wartość maksymalna	0,201134
Wartość minimalna	0,016772
Wartość średnia *m*	0,122
Odchylenie standardowe σ	0,045472
Współczynnik zmienności	0,371583
Mediana	0,120728
Kurtoza	-0,140458099
Skośność	-0,368422971

Źródło: Jak w tabeli 1.

Wielkości zużycia kleju poliuretanowego wykazują słabą koncentrację wokół średniej oraz prawostronną asymetrię. Na podstawie wielkości zużycia kleju (w kg/t) w ostatnich trzech latach stwierdzono, że wielkość zużycia jest zmienną losową o rozkładzie normalnym $N(0{,}122;\ 0{,}045472)$ (badanie wykonano testem Kołmogorowa-Smirnowa na poziomie istotności 0,05). Można także zauważyć, że oba materiały wykazują dużą zmienność zużycia.

Dodatkowo stwierdzono na poziomie istotności 0,05, że nie występuje korelacja liniowa pomiędzy wielkościami zużycia drewna i kleju poliuretanowego.

W tabeli 3 przedstawiono średnie ceny jednostkowe drewna, kleju, planowane roczne wydobycie oraz planowaną roczną kwotę wydatków na zakup drewna i kleju poliuretanowego.

Tabela 3

Ceny jednostkowe drewna, kleju, planowane roczne wydobycie oraz planowana roczna kwota wydatków na drewno i klej poliuretanowy

Cena za kg kleju – c_1	13,75 zł
Cena za m^3 drewna – c_2	296 zł
Planowane wydobycie – W	4 000 000 t
Maksymalny koszt zakupu materiałów – K	10 970 000,00 zł

Źródło: Jak w tabeli 1.

W modelu (4) zastosowano liniowe wagi w_t (t = 1, 2,…,36) i rozwiązano zagadnienie dla różnych wartości parametrów sterujących.

W tabeli 4 zamieszczono wartości rozwiązań optymalnych zagadnienia nieliniowego (4) przy różnych wartościach parametrów sterujących (wag). Dodatkowo w tabeli zamieszczono wielkości wskaźników nadmiaru i braku materiału dla kleju poliuretanowego i drewna kopalnianego.

Tabela 4

Wartości rozwiązań zagadnienia (4) dla różnych wartości parametrów sterujących

Wagi	kryteriów	Wagi materiałów		z_1	z_2	W_{sb1}	W_{sb2}	$F(z_1)$	$F(z_1)$	Wielkość zamówienia – klej (kg)	Wielkość zamówienia – drewno (m^3)	Koszt całkowity zamówienia
α_a	0	u_1(klej)	0,5	0,12202	0,00360	1,000	1,250	0,497	0,508	488082,6	14388,1	10 970 000
α_c	1	u_2(drewno)	0,5									
α_a	1	u_1(klej)	0,5	0,12202	0,00360	1,000	1,250	0,497	0,508	488082,6	14388,1	10 970 000
α_c	0	u_2(drewno)	0,5									
α_a	0,5	u_1(klej)	0,5	0,12466	0,00347	1,161	1,000	0,520	0,451	498634,6	13897,9	10 970 000
α_c	0,5	u_2(drewno)	0,5									
α_a	0,5	u_1(klej)	0,6	0,12466	0,00347	1,161	1,000	0,520	0,451	498634,6	13897,9	10 970 000
α_c	0,5	u_2(drewno)	0,4									
α_a	0,5	u_1(klej)	0,9	0,12466	0,00347	1,161	1,000	0,520	0,451	498633,6	13897,9	10 970 000
α_c	0,5	u_2(drewno)	0,1									
α_a	0,5	u_1(klej)	0,1	0,12202	0,00360	1,000	1,250	0,497	0,508	488082,5	14388,1	10 970 000
α_c	0,5	u_2(drewno)	0,9									
α_a	0,7	u_1(klej)	0,6	0,12466	0,00347	1,161	1,000	0,520	0,451	498634,6	13897,9	10 970 000
α_c	0,3	u_2(drewno)	0,4									
α_a	0,5	u_1(klej)	0,4	0,12202	0,00360	1,000	1,250	0,497	0,508	488082,6	14388,1	10 970 000
α_c	0,5	u_2(drewno)	0,6									
α_a	0,7	u_1(klej)	0,4	0,12202	0,00360	1,000	1,250	0,497	0,508	488082,5	14388,1	10 970 000
α_c	0,3	u_2(drewno)	0,6									
α_a	0,7	u_1(klej)	0,6	0,12466	0,00347	1,161	1,000	0,520	0,451	498634,6	13897,9	10 970 000
α_c	0,3	u_2(drewno)	0,4									
α_a	0,8	u_1(klej)	0,6	0,12466	0,00347	1,161	1,000	0,520	0,451	498634,1	13897,9	10 970 000
α_c	0,2	u_2(drewno)	0,4									
α_a	0,2	u_1(klej)	0,4	0,12202	0,00360	1,000	1,250	0,497	0,508	488082,6	14388,1	10 970 000
α_c	0,8	u_2(drewno)	0,6									

Źródło: Jak w tabeli 1.

Jak widać, dla różnych parametrów sterujących otrzymano rozwiązania niewiele się od siebie różniące. Jeśli przyjmie się zamówienie drewna kopalnianego $z_2 = 0{,}0036$ m^3/t, to przy założeniu, że warunki w kopalni się nie zmienią, będą się tworzyć zapasy drewna ($W_{sb2} = 1{,}25$). Jeśli przyjmie się zamówienie kleju poliuretanowego $z_1 = 0{,}12466$ kg/t, to przy założeniu, że warunki w kopalni się nie zmienią, będą się tworzyć zapasy kleju ($W_{sb1} = 1{,}16$). Opierając się na powyższych rozważaniach, decydent musi wybrać rozwiązanie, które może być podstawą ustalania planów asortymentowych w poszczególnych grupach przetargowych drewna kopalnianego oraz kleju poliuretanowego.

Podsumowanie

Na podstawie analizy przykładu można stwierdzić, że wielkość rocznego zamówienia zależy od wartości parametrów sterujących. Stąd zaproponowana metoda powinna być stosowana w postaci interaktywnej, w której decydent będzie określał wielkości parametrów sterujących, porównując koszty zamówień oraz wielkości wskaźników W_{sb}. Zaproponowany model może służyć jako pomoc przy planowaniu rocznych zamówień na materiały do produkcji.

Literatura

1. Ameljańczyk A.: *Optymalizacja wielokryterialna w problemach sterowania i zarządzania*. Ossolineum, Wrocław 1984.
2. Domański Cz. (red): *Nieklasyczne metody oceny efektywności i ryzyka*. PWE, Warszawa 2011.
3. Jakowska-Suwalska K., Wolny M., Sojda A.: *Wielokryterialny model sterowania zapasami*. ZN Politechniki Śląskiej, seria „Organizacja i Zarządzanie" 2011, 57, s. 169-182.
4. Jakowska-Suwalska K., Wolny M., Sojda A.: *Wielokryterialne sterowanie zapasami jako element wspomagania potrzeb materiałowych***.** „Zarządzanie i Edukacja" 2011, 96, s. 271-280.
5. Konarzewska-Gubała E.: *Programowanie przy wielorakości celów*. PWN, Warszawa 1980.
6. Nowak M.: *Interaktywne wielokryterialne wspomaganie decyzji w warunkach ryzyka. Metody i zastosowania*. Akademia Ekonomiczna, Katowice 2008.
7. Kukuła K.: *Metoda unitaryzacji zerowej*. Wydawnictwo Naukowe PWN, Warszawa 2000.
8. Krzyżaniak S., Cyplik P.: *Zapasy i magazynowanie*. Biblioteka Logistyka, Poznań 2007.
9. Ogryczak W.: *Wielokryterialna optymalizacja liniowa i dyskretna.* Wydawnictwa Uniwersytetu Warszawskiego, Warszawa 1997.

10. Prusek S., Stałęga S., Stochel D.: *Metody i środki przeznaczone do uszczelniania i wzmacniania górotworu oraz obudowy wyrobisk*. Prace Naukowe Głównego Instytutu Górnictwa nr 863, 2005.

11. Sarjusz-Wolski Z.: *Sterowanie zapasami w przedsiębiorstwie*. PWE, Warszawa 2000.

THE MULTI-CRITERIA, NONLINEAR MODEL OF ORDER SIZE OF MATERIALS FOR HARD-COAL MINE

Summary

In the work there is a multi-criteria model of order size of materials used in production presented. It was assumed that the consumption size of each material is a random variable of known probability distribution. In the model with the purchase cost of materials ordered limited there were three criteria: order size, probability of lack of materials in the production process, deviations of order size from the consumption size in the past periods.

It was shown on an example how to use the model for determining the order sizes for polyurethane adhesive and wood in one of the hard-coal mines.

Anna Janiga-Ćmiel

WYZNACZENIE OKRESU RÓWNOWAGI I STABILIZACJI DŁUGOOKRESOWEJ

Wprowadzenie

W rozwoju każdego zjawiska niezależnie od tego, jak rozwój ten jest ukształtowany przez trend i wahania, można wyznaczyć okres równowagi istotny dla krótkiego lub długiego okresu i okres ustabilizowanej zależności, również istotny dla krótkiego lub długiego okresu.

W okresie równowagi są zachowane parametry rozkładu charakteryzujące wartość oczekiwaną, wariancję i asymetrię, ponieważ w okresie tym ma się do czynienia z niezmienniczością rozkładu badanych zmiennych. Z kolei okres ustabilizowanej zależności dotyczy kolejnych lat, w których nie obserwuje się istotnych zmian w charakterze, sile i kierunku zależności. Długookresowa równowaga dotyczy okresu, w którym istnieją mechanizmy samoregulujące pozwalające osiągnąć stan oczekiwany. Jedna z metod wyznaczania przedmiotowych podokresów opiera się na analizie przyrostów badanych zmiennych.

Zarówno w przypadku gospodarki Polski, jak i gospodarek państw Unii Europejskiej zostanie wyznaczony najdłuższy możliwy okres ustabilizowanego rozwoju, czyli taki, w którym główne wskaźniki nie wykażą istotnych zmian. Taki okres jest potrzebny w celu dokonania porównania rozwoju gospodarki w Polsce i w krajach Unii Europejskiej.

Rozwój gospodarczy można scharakteryzować w dwojaki sposób. Jednym z nich jest analiza wielowymiarowa wybranego, możliwie najliczniejszego zbioru czynników. Mogą być one stymulantami lub destymulantami rozwoju gospodarczego. Podstawą drugiego ze sposobów jest zmienna charakteryzująca PKB ujmująca w pewnym sensie w sposób syntetyczny czynniki wykorzystane w pierwszym sposobie. W niniejszym artykule posłużono się drugą metodą wykorzystującą PKB. Ponadto analizie poddano gospodarki następujących krajów: Polski, Francji, Wielkiej Brytanii, Belgii, Holandii. Analizę wykonano na podstawie danych obejmujących lata 1949-2006 zaczerpniętych z Roczników Statystycznych i sprowadzonych uprzednio do postaci wzajemnie porównywalnych.

1. Wyznaczenie macierzy mnożników długookresowych

Badanie podobieństwa stanu gospodarki w różnych okresach oparto na macierzy wartości mnożników charakteryzujących wariancje i kowariancje stanów rozwoju gospodarek w poszczególnych krajach. Liczba obserwacji wyjściowego szeregu powinna obejmować długi okres. Wymiar szeregu czasowego odpowiada liczbie porównywanych krajów. W dalszej analizie przez *m* oznaczono ilość badanych krajów Unii Europejskiej z wyłączeniem Polski. Macierz danych empirycznych szeregu czasowego wielowymiarowego oznaczono przez *W*. Macierz *W* jest wielowymiarowym szeregiem czasowym jednostkowego PKB w rozpatrywanych krajach. Macierz *W* przedstawiono w tabeli 1.

Tabela 1

Wskaźniki poziomu jednostkowego PKB Polski i krajów UE

t	Lata	**Polska**	**Francja**	**Wielka Brytania**	**Holandia**	**Belgia**
1	2	3	4	5	6	7
1	1958	0,0048	0,0099	0,0420	0,0056	0,0134
2	1959	0,0049	0,0157	0,0363	0,0033	0,0135
3	1960	0,0049	0,0185	0,0405	0,0041	0,0143
4	1961	0,0050	0,0208	0,0451	0,0048	0,0180
5	1962	0,0054	0,0160	0,0501	0,0082	0,0184
6	1963	0,0054	0,0187	0,0555	0,0085	0,0190
7	1964	0,0054	0,0291	0,0614	0,0082	0,0190
8	1965	0,0055	0,0327	0,0597	0,0115	0,0190
9	1966	0,0055	0,0378	0,0598	0,0122	0,0200
10	1967	0,0053	0,0464	0,0706	0,0178	0,0222
11	1968	0,0053	0,0487	0,0734	0,0213	0,0257
12	1969	0,0054	0,0503	0,0763	0,0236	0,0284
13	1970	0,0056	0,0525	0,0809	0,0255	0,0301
14	1971	0,0063	0,0417	0,0636	0,0206	0,0325
15	1972	0,0098	0,0573	0,0938	0,0220	0,0341
16	1973	0,0100	0,0607	0,1180	0,0234	0,0363
17	1974	0,0102	0,0642	0,1273	0,0248	0,0370
18	1975	0,0103	0,0751	0,1346	0,0261	0,0374
19	1976	0,0106	0,0781	0,1386	0,0324	0,0483
20	1977	0,0106	0,0798	0,1416	0,0365	0,0532
21	1978	0,0109	0,0814	0,1443	0,0406	0,0579
22	1979	0,0111	0,0838	0,1474	0,0461	0,0681
23	1980	0,0107	0,0858	0,1535	0,0505	0,0783
24	1981	0,0107	0,0887	0,1622	0,0555	0,0866
25	1982	0,0094	0,0921	0,1644	0,0566	0,0966
26	1983	0,0135	0,0956	0,1710	0,0625	0,1068
27	1984	0,0137	0,1000	0,1833	0,0682	0,1134

cd. tabeli 1

1	2	3	4	5	6	7
28	1985	0,0140	0,1050	0,1890	0,0732	0,1205
29	1986	0,0139	0,1098	0,1949	0,0787	0,1283
30	1987	0,0143	0,1151	0,2036	0,0842	0,1287
31	1988	0,0222	0,1221	0,2257	0,0877	0,1398
32	1989	0,0198	0,1279	0,2163	0,0915	0,1483
33	1990	0,0182	0,1324	0,2400	0,0933	0,1636
34	1991	0,0189	0,1494	0,2315	0,0975	0,1922
35	1992	0,0186	0,1436	0,2420	0,1149	0,1809
36	1993	0,0189	0,1470	0,2185	0,1052	0,1878
37	1994	0,0369	0,1655	0,2486	0,1049	0,2225
38	1995	0,0380	0,1847	0,2556	0,1217	0,2342
39	1996	0,0391	0,1870	0,2705	0,1181	0,2336
40	1997	0,0403	0,1900	0,2772	0,1369	0,2370
41	1998	0,0421	0,1929	0,2838	0,1286	0,2384
42	1999	0,0430	0,1965	0,2904	0,1296	0,2368
43	2000	0,0529	0,2008	0,2918	0,1478	0,2360
44	2001	0,0683	0,2055	0,3128	0,1504	0,2346
45	2002	0,0690	0,2101	0,2883	0,1734	0,2343
46	2003	0,0809	0,2157	0,2881	0,1691	0,2345
47	2004	0,0974	0,2217	0,3052	0,1757	0,2362
48	2005	0,1090	0,2266	0,3090	0,1983	0,2390
49	2006	0,1303	0,2372	0,3156	0,2060	0,2429

Na podstawie macierzy W wyznaczono macierz teoretyczną przybliżonych wartości mnożników długookresowych [5]. Macierz mnożników oznaczono przez π, przy czym:

$$\pi = \frac{1}{p} W^T W \tag{1}$$

gdzie p oznacza ilość okresów, jakich dotyczą szeregi czasowe. Jest to macierz w postaci:

$$\pi = \begin{bmatrix} S_{00} & S_{01} & S_{02} & S_{03} & S_{04} \\ S_{10} & \ddots & & & \\ S_{20} & & \pi_{mm} & & \\ S_{30} & & & \ddots & \\ S_{40} & & & & \ddots \end{bmatrix} = \begin{bmatrix} S_{00} & \pi_0^T \\ \pi_0 & \pi_{mm} \end{bmatrix} \tag{2}$$

Dodatkowo przez π_0 oznaczono wektor:

$$\pi_0 = \begin{bmatrix} S_{10} \\ S_{20} \\ S_{30} \\ S_{40} \end{bmatrix} \tag{3}$$

Elementami macierzy są iloczyny skalarne w postaci:

$$s_{ij} = \frac{1}{p} \sum_{t=1}^{p} w_{it}^T w_{jt} \tag{4}$$

Dla *i, j = 0,…,m*, przy czym 0 oznacza wektor wskaźnika rozwoju gospodarczego w Polsce, a indeksy j = 1,…,m dotyczą wskaźników rozwoju gospodarczego w rozpatrywanych krajach. Przez B oznaczono podmacierz macierzy *W* dotyczącą krajów innych niż Polska. Przez π_{mm} oznaczono macierz powstałą z macierzy π przez skreślenie pierwszego wiersza i pierwszej kolumny.

2. Wyznaczenie macierzy wag

Poszczególne wariancje i kowariancje zawarte w macierzy π można odpowiednio zrangować przez przyporządkowanie im macierzy wag. Każda z wag będzie ilustrowała relację między stanem gospodarki w kraju *i* oraz stanem gospodarki w kraju *j*. Wyznaczoną macierz wag oznaczono przez *K*. Macierze *K* oraz *B* mają ten sam wymiar, tzn. mają po *p* wierszy i *m* kolumn.

Macierz wag [5] jest zdefiniowana następująco:

$$K = S_{00}(B\pi_{mm}B^T)^{-1}B \tag{5}$$

Jest tu rozpatrywany model jednorównaniowy, więc w powyższej definicji S_{00} jest macierzą jednoelementową określoną wzorem:

$$S_{00} = \frac{1}{p} \sum_{t=1}^{N} w_{0t}^T w_{0t} \tag{6}$$

Wartość S_{00} to czynnik stanowiący charakterystykę rozpatrywanego rozwoju gospodarczego w Polsce.

Uwzględniając przedstawione oznaczenia, otrzymano macierz struktury zrównoważonego rozwoju, która ma postać:

$$S = B \otimes K \tag{7}$$

Przedstawiony iloczyn $\otimes$ macierzy B oraz K jest rozumiany jako macierz iloczynów skalarnych odpowiednich wektorów kolumn. Oznaczając elementy macierzy B przez $\lfloor w_{ij} \rfloor$, macierzy K przez $\lfloor k_{ij} \rfloor$, buduje się macierz S elementów w postaci:

$$s_{ij} = w_{ij} \cdot k_{ij} \tag{8}$$

Macierze B i K są jednakowych wymiarów i takie same wymiary ma macierz S, która jest macierzą iloczynów elementów na tych samych pozycjach w macierzach B i K.

Analiza macierzy S pozwala na wykrycie kolejnych podokresów z przedziału p okresowego, w którym współczynnik korelacji wyważonych kolumn PKB będzie najwyższy. Maksymalnej długości przedział wyznaczony w ten sposób będzie okresem równowagi i stałej zależności długookresowej w przedziale danych historycznych. Przedstawiona wyżej macierz S stanowi punkt wyjścia do wyznaczenia dalszych macierzy charakteryzujących relację między dynamiką rozwoju gospodarczego w poszczególnych państwach. Macierze te przedstawiają zasadniczą charakterystykę wielowymiarowego rozwoju i są oznaczone przez E i F. Macierz F dotyczy rozwoju gospodarczego w wybranych krajach, macierz E przedstawia wyważone różnice rozwoju gospodarczego w Polsce i w innych krajach Unii Europejskiej uwzględnionych łącznie. Macierz F jest wyznaczona zgodnie ze wzorem:

$$F = \frac{1}{p}(B \otimes K)^T (B \otimes K) = \frac{1}{p} S^T S \tag{9}$$

Macierz E wyznaczono według wzoru:

$$E = F - \pi_0 S_{00}^{-1} \pi_0^T \tag{10}$$

Macierze E i F są macierzami kwadratowymi o wymiarze *mxm,* gdzie m to ilość krajów, z którymi porównuje się rozwój gospodarczy w Polsce.

Dla macierzy F, E wyznacza się wartości własne. O zrównoważonym rozwoju zjawiska w porównywanych zbiorowościach można mówić wtedy, gdy macierze E i F (zgodnie z [2]) są jednakowe lub nie wykazują istotnej statystycznie różnicy. Ponadto jeśli w uporządkowanych ciągach wartości własnych stwierdza się różnice i dla kolejnych wartości własnych te różnice będą coraz to niższe, to ma się do czynienia ze zjawiskiem, które w określonych zbiorowościach dąży do równowagi. Natomiast jeżeli przyrosty względne kolejnych wartości własnych są coraz to większe, oznacza to, że zjawisko w badanych zbiorowościach nigdy nie osiągnie wzajemnej równowagi. Minimalny okres równowagi w zakresie danych historycznych można wyznaczyć jako:

$$t_{\min} = \lambda_1 \cdot \lambda_2 \cdot \lambda_3 \cdot \lambda_4 \cdot p \tag{11}$$

gdzie mnożnik jest iloczynem wartości własnych λ_i macierzy *E*. Liczba wartości własnych odpowiada liczebności grupy krajów, z którymi łącznie porównuje się rozwój gospodarczy Polski.

3. Wyznaczenie macierzy struktury zrównoważonego rozwoju

Zgodnie z rozważaniami teoretycznymi analizę rozpoczyna się od wyznaczenia macierzy mnożników długookresowych π:

Tabela 2

Macierz mnożników długookresowych π

	Polska	Francja	Wielka Brytania	Holandia	Belgia
Polska	0,00146	0,00435	0,00632	0,00328	0,00482
Francja	0,00435	0,01613	0,02443	0,01153	0,01807
Wielka Brytania	0,00632	0,02443	0,03752	0,01727	0,02725
Holandia	0,00328	0,01153	0,01727	0,00844	0,01302
Belgia	0,00482	0,01807	0,02725	0,01302	0,02068

Na podstawie analizy otrzymanych mnożników można stwierdzić, że w Polsce słonność do długookresowej równowagi była najniższa, natomiast w krajach Unii Europejskiej znacząco wyższa.

W tabelach 3 i 4 przestawiono charakterystykę podążania poszczególnych gospodarek do stanu równowagi. Widać, że w okresie ostatnich piętnastu lat wskaźniki te są odpowiednio niższe dla wybranych krajów Unii Europejskiej z wyjątkiem Holandii, co oznacza, że dla tych krajów istotne znaczenie ma równowaga rozwoju gospodarczego w okresie całego rozpatrywanego pięćdziesięciolecia. Wskaźnik dla Polski dla okresu pięćdziesięciu lat wynosi 7,22%, a dla ostatnich piętnastu lat wzrasta do 13,43%, co oznacza istotność równowagi w okresie piętnastu ostatnich lat.

Tabela 3

Udział mnożnika dla rozpatrywanego okresu

Kraje	Udział mnożnika
Polska	7,22%
Francja	21,48%
Wielka Brytania	31,24%
Holandia	16,21%
Belgia	23,85%

Spośród badanych krajów najwyższą skłonność do długookresowej równowagi wykazuje Wielka Brytania w skali 31,24% całkowitej zmienności w rozwoju gospodarki. Belgia i Francja charakteryzowały się podobnym poziomem, ponad 20%, Holandia 16%, Polska 7%. Wynik ten otrzymano z uwzględnieniem czterdziestu dziewięciu lat rozwoju gospodarczego tych krajów. Powtarzając tę samą analizę dla ostatnich piętnastu lat rozpatrywanego okresu, można stwierdzić istotne zmiany w dążeniu do równowagi:

Tabela 4

Udział mnożnika dla ostatnich piętnastu lat

Kraje	Udział mnożnika
Polska	13,43%
Francja	25,44%
Wielka Brytania	35,33%
Holandia	20,12%
Belgia	25,67%

Widać, że dążność do równowagi rozwoju gospodarczego w krótszym okresie jest wyższa, a spośród badanych krajów najwyższy wzrost notuje Polska. W celu dalszego i pogłębionego porównania rozwoju gospodarczego Polski i państw Unii Europejskiej wyznaczono macierz *F* zgodnie ze wzorem:

$$F = \frac{1}{p}(W \otimes K)^T (W \otimes K) = \frac{1}{p} S^T S \tag{12}$$

Jest to macierz w postaci:

Tabela 5

Mnożniki zrównoważonego rozwoju w UE

	Francja	Wielka Brytania	Holandia	Belgia
Francja	2,9	-1,9	-0,7	-0,2
Wielka Brytania	-1,9	1,3	0,5	0,1
Holandia	-0,7	0,5	0,2	0,1
Belgia	-0,2	0,1	0,1	0

Macierz ta ilustruje mnożniki zrównoważonego rozwoju gospodarki w państwach Unii Europejskiej. W dalszej kolejności wyznaczono macierz *E* zgodnie ze wzorem:

$$E = F - \pi_0 S_{00}^{-1} \pi_0^T \tag{13}$$

Macierz *E* przedstawia mnożniki zrównoważonego rozwoju gospodarki z wyłączeniem wpływu Polski na rozwój gospodarki w tych krajach i odwrotnie, otrzymując tym samym macierz w postaci:

Tabela 6

Mnożniki zrównoważonego rozwoju gospodarczego krajów UE z wyłączeniem Polski

	Francja	Wielka Brytania	Holandia	Belgia
Francja	2,92	-1,93	-0,739	-0,239
Wielka Brytania	-1,93	1,291	0,443	0,125
Holandia	-0,739	0,443	0,2	0,045
Belgia	-0,239	0,125	0,045	0,005

Jednak dla nas interesujący jest iloczyn wartości własnych jednej macierzy i drugiej; iloczyn wartości własnych jest równy wartości wyznacznika odpowiedniej macierzy:

$$\lambda_1 \cdot ..... \cdot \lambda_4 = \det F = 0{,}00001708 \tag{14}$$

Natomiast dla macierzy *E* otrzymano:

$$\lambda_1 \cdot ..... \cdot \lambda_4 = \det E = 0{,}00010795 \tag{15}$$

Przedstawione iloczyny wartości własnych nie mają zasadniczo interpretacji ekonomicznej. Interpretacji podlega jedynie ich zmiana. Przedstawione wartości wyznaczników spełniają następujący związek:

$$\det E = 6{,}3205 \cdot \det F \tag{16}$$

Oznacza to, że wyznacznik macierzy, w której pominięto powiązania rozwoju gospodarek Unii Europejskiej z rozwojem gospodarki w Polsce, zwiększył się 6,3205 razy, co oznacza brak jakiejkolwiek współzależności między rozwojem gospodarki w Polsce w okresie rozpatrywanych czterdziestu dziewięciu lat. Współzależność rozwoju wystepowałaby w przypadku, gdyby wartości *det*E i *det*F istotnie się nie różniły.

W dalszym toku analizy z okresu czterdziestu dziewięciu rozpatrywanych lat wyodrębniono krótszy podokres obejmujący piętnaście lat, tzn. lata dziewięćdziesiąte i po 2000 roku. Dla rozpatrywanych piętnastu lat powtórzono analizę równowagi rozwoju gospodarki Polski i wybranych krajów Unii Europejskiej. Macierz *F* przyjęła postać:

Tabela 7

Mnożniki zrównoważonego rozwoju w UE dla piętnastu lat

	Francja	Wielka Brytania	Holandia	Belgia
Francja	-1936,67	-983,996	1154,162	1738,798
Wielka Brytania	-304,538	-92,003	167,692	233,593
Holandia	992,001	419,411	-569,027	-829,409
Belgia	1236,342	658,355	-744,711	-1136,102

Analogiczna macierz *E* przyjęła następującą postać:

Tabela 8

Mnożniki zrównoważonego rozwoju gospodarek krajów UE z wyłączeniem Polski dla piętnastu lat

	Francja	Wielka Brytania	Holandia	Belgia
Francja	-1936,704	-984,043	1154,136	1738,761
Wielka Brytania	-304,585	-92,069	167,655	233,54
Holandia	991,975	419,374	-569,048	-829,438
Belgia	1236,305	658,302	-744,74	-1136,145

Okazuje się, że w rozpatrywanym przypadku iloczyny wartości własnych również się zmieniły, przyjmując wartość dla macierzy *F* odpowiednio:

$$\lambda_1 \cdot ..... \cdot \lambda_4 = \det F = 1817338{,}421 \tag{17}$$

Natomiast dla macierzy *E* otrzymano:

$$\lambda_1 \cdot ..... \cdot \lambda_4 = \det E = 1712559{,}000 \tag{18}$$

Relacja między wyznaczonymi wartościami wyznaczników jest następująca:

$$\frac{\det E}{\det F} = 0{,}9424 \tag{19}$$

Powyższa zależność oznacza, że pomiędzy wartościami wyznaczników macierzy *E* oraz *F* nie występują istotne statystycznie różnice. Z ekonomicznego punktu widzenia należy zaznaczyć, że rozwój gospodarki polskiej i rozwój gospodarek państw Unii Europejskiej wraz z upływem czasu staje się coraz bardziej do siebie zbliżony. Nie można jeszcze mówić o pełnej równowadze rozwoju gospodarek w Polsce i w państwach Unii Europejskiej.

Podsumowanie

Przedstawiona w artykule metoda Johansena wyznaczania okresu równowagi i stabilizacji długookresowej pozwala na ocenę porównawczą dynamiki szeregów czasowych. Można ją rozpatrywać w całym okresie danych historycznych lub poszczególnych podokresach. Zastosowana metoda pozwala wyznaczyć minimalny okres równowagi lub odpowiedzieć na pytanie, w jakim okresie procentowym rozwój gospodarczy w rozpatrywanym czasie można uznać za ustabilizowany. Dla całego pięćdziesięciolecia otrzymano niskie wartości własne, a dla ostatnich piętnastu lat istotnie wyższe. Oznacza to, że w ostatnim piętnastoleciu występuje znacząca kointegracja rozpatrywanych gospodarek.

Literatura

1. Grabowski W., Welfe A.: *Ekonometria. Zbiór zadań*. PWE, Warszawa 2010.
2. Johansen S.: *Likelihood Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford 1996.
3. Rocznik Statystyczny GUS. Warszawa 1958-2006.
4. Welfe A., Karp P., Kębłowski P.: *Mechanizmy makroekonomiczne w gospodarce polskiej. Analiza ekonometryczna*. WUŁ, Łódź 2006.
5. Welfe A.: *Gospodarka Polski w okresie transformacji. Zasady modelowania ekonometrycznego*. PWE, Warszawa 2000.

DETERMINATION OF THE PERIOD OF LONG-TERM EQUILIBRIUM AND STABILITY

Summary

The study examines the development of the Polish economy as well as the economies of selected European Union countries in the period from 1949 to 2006. Much space is devoted to a comparative analysis of the development economies in the countries concerned. Based on statistical data appropriate synthetic variables were set. Much space is devoted to the theory of the Johansen's method, to present the dependencies occurring in the dynamics of economic development in subsequent subperiods. The method allows for a comparative assessment of the dynamics of time series. The methods are adopted to examine the level of economic development, to determine the period of long-term equilibrium and stability.

Adrianna Mastalerz-Kodzis

TEORIA FAL ELLIOTTA A TEORIA FRAKTALI – PODOBIEŃSTWA I RÓŻNICE W PODEJŚCIU DO MODELOWANIA SZEREGÓW ORAZ OPISU ZACHOWAŃ INWESTORA

Wprowadzenie

Opisem zachowania inwestora w procesie podejmowania decyzji zajmuje się wiele teorii ekonomicznych oraz psychologicznych. Niniejszy artykuł wskazuje na dwie z nich: teorię fal Elliotta oraz teorię bazującą na geometrii fraktalnej. Obydwie teorie są interesujące z matematycznego, a w szczególności geometrycznego punktu widzenia, i uwzględniają w procesie podejmowania decyzji inwestycyjnych wpływ danych historycznych na zachowania inwestora, jednakże opierają się na różnych przesłankach, wykorzystują inne narzędzia matematyczno-ekonometryczne. Zaskakujące są jednak pewne podobieństwa w modelowaniu np. szeregów czasowych za pomocą tych dwóch podejść. Niniejszy artykuł jest zatem polemiką na temat możliwości konstrukcji modelu zachowania inwestora z zastosowaniem wybranych metod.

Artykuł składa się z trzech części. W pierwszej krótko podano opis teorii fal, w drugiej skupiono się na wybranych elementach teorii fraktali, zaś cześć trzecia ma charakter empiryczny.

1. Elementy teorii fal

Teoria fal została stworzona w latach 30. XX wieku przez Ralpha Nelsona Elliotta, który próbował uwzględnić aspekt psychologiczny w zachowaniu uczestników rynku [4]. Teoria fal Elliotta jest analizą zmian na rynku akcji w kontekście psychologii tłumu. Między innymi na bazie teorii Elliotta rozwinęła się analiza techniczna rynków finansowych. Połączenie metod analizy technicznej, fundamentalnej oraz teorii fraktali zdaje się być ciekawym uzupełnieniem dotychczas stosowanych metod w teorii inwestowania na giełdzie.

1.1. Cykle, ciągi liczb Fibonacciego

Istnieje teoria głosząca, że wszystkie zjawiska na świecie mają charakter cykliczny [6]. Niektóre cykle trwają lata, inne miesiące, tygodnie, dni czy minuty. Rynki finansowe także zachowują się cyklicznie.

Zgodnie z teorią Elliotta istnieje dziewięć stopni trendu ze względu na długość cyklu:
- wielki supercykl – od kilku dziesięcioleci do kilku wieków,
- supercykl – od kilku lat do kilku dziesięcioleci,
- cykl – od roku do kilku lat,
- fale pierwotne – od kilku miesięcy do kilku lat,
- fala pośrednie – od kilku tygodni do kilku miesięcy,
- fale mniejsze,
- fale minutowe,
- fale minutkowe,
- fale subminutkowe.

Ze względu na brak danych trudno jest rozważać wielki supercykl. Jednak efekty supercykli aż do fal minutowych można obserwować biorąc pod uwagę dane makroekonomiczne oraz giełdowe (przykłady w rozdziale 3 niniejszego artykułu).

Teoria Elliotta opiera się na analizie trzech elementów rynku: kształtu fali, proporcji i czasu. Uwzględnia się w niej aspekty psychologiczne, które oddziałują na rynki finansowe. Podstawowe założenia tej teorii to:
- rynek podlega cyklicznemu zachowaniu pięciofalowego ruchu w górę oraz trójfalowego spadku w dół,
- każdej akcji odpowiada reakcja.

Według teorii Elliotta ceny zmieniają się zgodnie z cyklami opartymi na ciągu liczb Fibonacciego. Może on być zdefiniowany rekurencyjnie w następujący sposób:

$$F_0 = 0,\ F_1 = 1,$$

$$F_n = F_{n\text{-}1} + F_{n\text{-}2},\ \text{dla } n \geq 2 \tag{1}$$

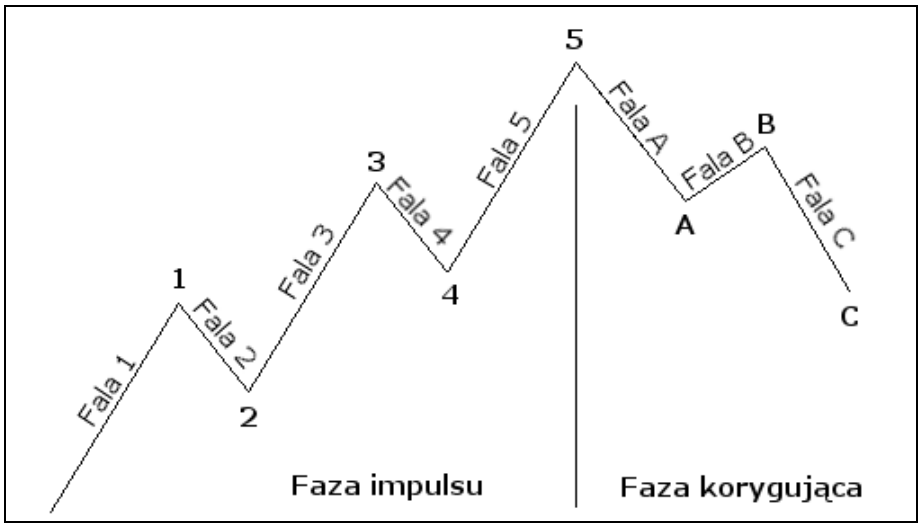
Ciąg ten posiada wiele interesujących własności. Granica ilorazu stosunku wyrazu ciągu do wyrazu go poprzedzającego wynosi 1,618. Liczba ta nazywana „złotym podziałem" jest jednym z rozwiązań równania:

$$\varphi^2 - \varphi - 1 = 0 \tag{2}$$

czyli:

$$\varphi = \frac{1+\sqrt{5}}{2} \tag{3}$$

Klasyczny wzrost składa się z pięciu fal, z których trzy stanowią fale wzrostu (1,3,5), zaś dwie kolejne fale korygujące (2,4). Podczas spadku powstają trzy fale: dwie są zniżkowe (A, C) i jedna wzrostowa (B).



Rys. 1. Pojedynczy cykl

Źródło: Opracowanie na podstawie [4].

1.2. Charakterystyka fal cyklu

Fale cyklu posiadają następujące cechy. Fala numer:

1. Jest początkiem nowego trendu, może zająć połowę długości trendu wzrostowego. Jest najczęściej najkrótszą z pięciu fal. W tym czasie rynek wykazuje się dużą dynamiką.
2. Zabiera cześć fali 1, nie może zejść poniżej poziomu początku fali 1.
3. Jest zazwyczaj najdłuższa, najszybciej zmieniają się ceny podczas jej trwania, występuje najwyższy wolumen obrotów, nigdy nie jest najkrótszą z fal. Coraz większą rolę odgrywają elementy fundamentalne rynku.
4. Koryguje wzrosty z fali 3. Nie powinna zachodzić na falę 1. Różni się nachyleniem od fali 2.
5. Wykazuje mniejszą dynamikę niż fala 3, w trakcie jej trwania dochodzi do spadków obrotów, osłabienia siły rynku i zmiany trendu.

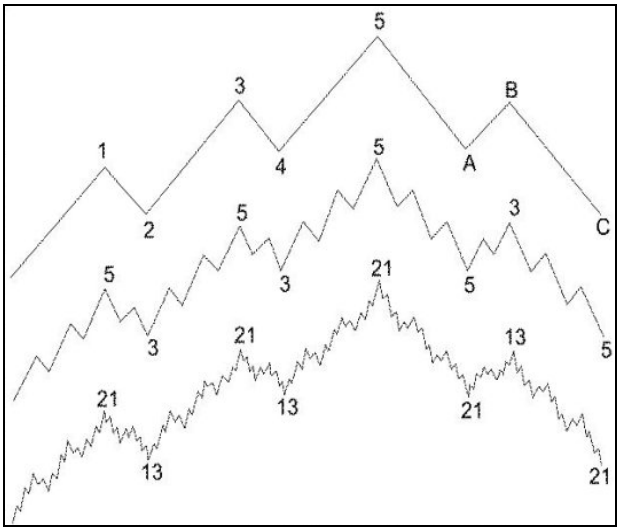
A. Jest początkiem trendu spadkowego. Często wraz ze spadkami cen dochodzi do wzrostu wolumenu, następuje korekta rynku.

B. Następuje odbicie do góry przy niskim wolumenie obrotów. Zwyżka może sięgnąć szczytu fali 5, jest widoczny podwójny szczyt, stanowi on sygnał zmiany trendu.
C. Kolejne zniżki. Ceny mogą spaść nawet poniżej dołka fali A.

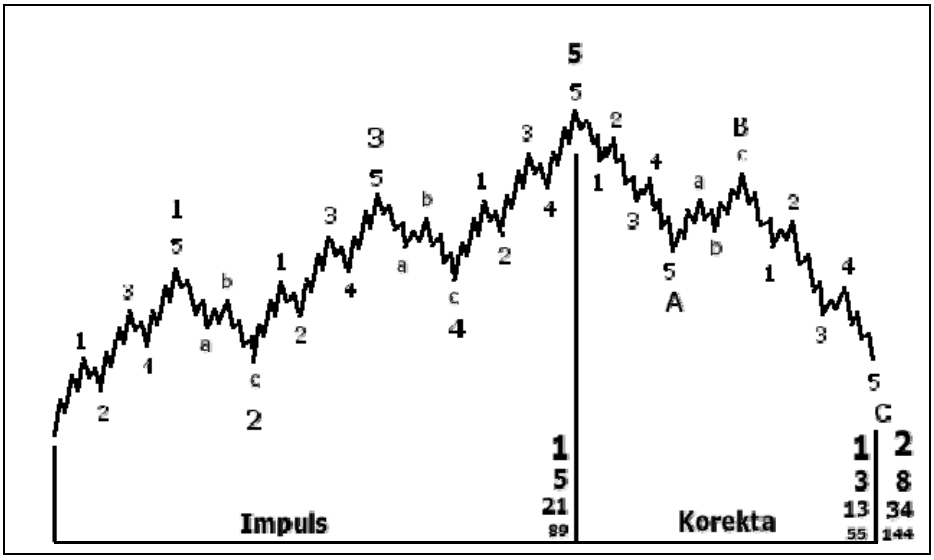
W praktyce fale mogą przybierać różne zmodyfikowane formy. Na przykład fale impulsu w trendzie wzrostowym mogą się wydłużać. Taka sytuacja zazwyczaj dotyczy jednej z fal impulsu, przy czym pozostałe dwie fale są do siebie podobne. Na przykład jeżeli wydłużaniu ulega fala 3, to wówczas fala 1 i 2 mają podobną długość, budowę i czas trwania. Zmianie mogą podlegać także fale korygujące.

Każda fala dzieli się na fale niższego stopnia, które dzielą się na fale jeszcze niższego stopnia. Na rysunku 2 przedstawiono dwie duże fale – wzrostową i korekcyjną, składające się na jeden pełny cykl. Dzielą się one na osiem mniejszych fal składających się z 34 fal jeszcze niższego stopnia. Przy kolejnym podziale uzyskano 144 fale jeszcze niższego stopnia.

Pełny cykl dzieli się na cykl hossy, w którym występują odpowiednio następujące ilości fal: 1, 5, 21, 89. W następującym po nim cyklu bessy można wyróżnić 1, 3, 13, 55 fal różnych stopni. Wszystkie wymienione powyżej liczby należą do ciągu Fibonacciego. Analizując fale korekty A, B, C, można dostrzec, że fale spadkowe A i C dzielą się na pięć fal. Wynika to z faktu, że zmierzają one w kierunku zgodnym z falą wyższego stopnia. Fala B składa się w tym przypadku z trzech fal, ponieważ jej kierunek jest przeciwny do fali wyższego stopnia.



a)



b)

Rys. 2. a) Różne poziomy fal Elliotta, b) Pełny cykl rynkowy (144 fale)

Źródło: [4].

1.3. Proporcje fal

Przy badaniu proporcji fal mają zastosowanie liczby z ciągu Fibonacciego. Główne zależności można zapisać w następujących punktach:

1. Jeżeli fala 3 się wydłuża, to minimalny jej zasięg stanowi sumę wartości poziomu dołka fali 2 oraz iloczynu długości fali 1 i liczby 1,618.
2. Zasięg fali 5 stanowi sumę wartości szczytu lub dna fali 1 oraz podwójnego iloczynu długości fali 1 i liczby 1,618.
3. Jeżeli fale 1 i 3 są jednakowej długości, wówczas najprawdopodobniej fala 5 będzie wydłużona. Jej zasięg oblicza się poprzez dodanie do poziomu dna fali 4 iloczynu liczby 1,618 oraz różnicy odległości szczytu fali 3 i poziomu dna fali 1.
4. W przypadku zygzaka (5-3-5) długość fali C oblicza się mnożąc liczbę 0,618 i długość fali A. Otrzymany wynik jest odejmowany od poziomu dna fali A.
5. Jeżeli ma miejsce płaska korekta (3-3-5), długość fali C stanowi 1,618 fali A.

Obliczone współczynniki procentowe Fibonacciego pozwalają określić wielkość przyszłej korekty cen. Wykorzystuje się zaokrąglone współczynniki 38%, 50%, 62%. Podczas silnego trendu korekta może wynieść 38%, natomiast w słabym trendzie – 62%.

Problem analizy czasu w teorii Elliotta jest uważany za najmniej istotny. W tym przypadku analitycy doszukują się również na osi czasu sekwencji liczb ciągu Fibonacciego.

Teoria Elliotta wygląda efektownie, budzi duże zainteresowanie wśród matematyków, a także inwestorów. Stosowanie analizy rynku z wykorzystaniem teorii fal wymaga dużego doświadczenia. Dlatego też równolegle do analizy z wykorzystaniem teorii Elliotta powinny być stosowane inne narzędzia i metody analityczne. Dopiero znalezienie w nich potwierdzenia przypuszczeń wysuniętych na podstawie teorii fal może prowadzić do właściwych wniosków.

Teoria fal Elliotta opiera się na regułach czysto matematycznych, graficznie fale w prostokątnym układzie współrzędnych są sumą odcinków (czasem o nieskończenie małej długości). Cykle długoletnie składają się z cykli krótszych, na które jeszcze składają się fale minutowe, minutkowe. Można zatem szukać analogii pomiędzy podziałem fal a elementami geometrii fraktalnej czy też multifraktalnej.

2. Wybrane elementy geometrii fraktalnej

Ceny akcji oraz indeksów giełdowych w wyniku gry rynkowej zmieniają się w czasie. Zmiany te są odnotowywane w dyskretnych momentach czasu (każdego dnia) bądź co minutę (notowania ciągłe). Wykresy obrazujące zachodzące zmiany w czasie cen instrumentów giełdowych są łamanymi o wymiarze topologicznym równym jeden.

Fraktalem nazywa się zbiór $A \subset \Re^n \quad (n \geq 1)$, który:

- ma wymiar Hausdorffa większy niż wymiar topologiczny*,
- posiada pewną formę samopodobieństwa (samoafiniczności), może to być samopodobieństwo przybliżone lub statystyczne,
- w miarę powiększania fragmentów zbioru A więcej jego detali staje się widocznych,
- jest zdefiniowany w prosty, często rekurencyjny sposób.

W modelach matematycznych stosowanych w finansach, modelach z czasem ciągłym stosuje się analizę stochastyczną. Wychodzi się z założenia, że dynamika procesu zmian cen instrumentów finansowych (akcji, indeksów giełdowych, kursów walutowych) jest opisana za pomocą stochastycznych równań różniczkowych Itô. Rozwiązaniami tych równań są procesy stochastyczne, których trajektorie aproksymują realne, quasi-fraktalne szeregi czasowe, tj. wykresy zmian cen instrumentów finansowych w czasie. Fraktalami losowymi są wykresy trajektorii procesów stochastycznych z czasem ciągłym, takie jak proces ru-

* Wymiar Hausdorffa można zastąpić wymiarem pojemnościowym, samopodobieństwa bądź też korelacyjnym. Zob. [7; 3].

chu Browna (standardowy, arytmetyczny, geometryczny), ułamkowy i multiułamkowy proces ruchu Browna [8; 10; 1].

Obiekty fraktalne posiadają własność samopodobieństwa. Można je podzielić na części o takiej własności, że każda z nich stanowi pomniejszoną kopię całości. Wykres wahań cen papierów wartościowych wygląda podobnie niezależnie od tego, czy się go pomniejszy, czy też powiększy, by dopasować go do określonej skali czasu lub cen. O samopodobieństwie mówi się wtedy, gdy każdy element obrazu zostaje pomniejszony lub powiększony w takim samym stosunku. Jednakże wykresy notowań na rynkach finansowych zachowują się „nie całkiem podobnie", biorąc pod uwagę wykres roczny, miesięczny bądź dzienny. Czasami oś czasu jest bardziej „ściśnięta" od osi cen, czasami na odwrót. Opisem tego typu zachowań rynku są multifraktale.

Procesy multifraktalne są zależne od funkcji Höldera $(H(t) \in (0,1))$, a ułamkowy ruch Browna zależy od wykładnika Hursta $(H \in (0,1))$.

Niech (X, d_x) i (Y, d_Y) będą przestrzeniami metrycznymi. Funkcję $f : X \to Y$ nazywamy funkcją Höldera o wykładniku α $(\alpha > 0)$, jeśli dla każdych $x, y \in X$ takich, że $d_X(x, y) < 1$ funkcja spełnia nierówność z dodatnią stałą c:

$$d_Y(f(x), f(y)) \le c \cdot d_X(x, y)^{\alpha} \tag{4}$$

Niech będzie dana funkcja $f : D \to \Re$ $(D \subset \Re)$ oraz parametr $\alpha \in (0,1)$. Funkcja $f : D \to \Re$ jest funkcją klasy C^{α} Höldera $(f \in C^{\alpha})$, jeżeli istnieją stałe $c > 0$ oraz $h_0 > 0$ takie, że dla każdego x oraz wszystkich h takich, że $0 < h \le h_0$ jest spełniona nierówność:

$$|f(x + h) - f(x)| \le c\, h^{\alpha} \tag{5}$$

Niech $x_0 \in D \subset \Re$. Funkcja $f : D \to \Re$ jest w punkcie x_0 funkcją klasy $C^{\alpha}_{x_0}$ Höldera, jeżeli istnieją stałe $\varepsilon, c > 0$ takie, że dla każdego $x \in (x_0 - \varepsilon, x_0 + \varepsilon)$ jest spełniona nierówność:

$$|f(x) - f(x_0)| \le c\, |x - x_0|^{\alpha} \tag{6}$$

Punktowym wykładnikiem Höldera funkcji f w punkcie x_0 nazywa się liczbę $\alpha_f(x_0)$ daną wzorem:

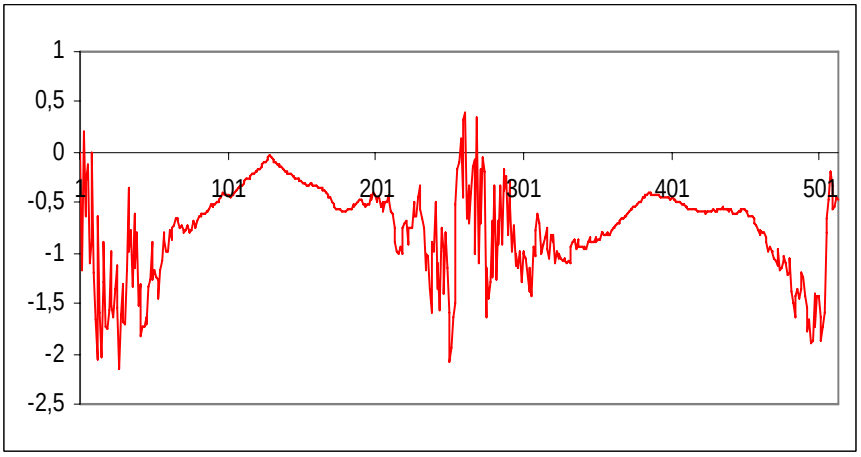
$$\alpha_f(x_0) = \sup\{\alpha : f \in C^{\alpha}_{x_0}\} \tag{7}$$

Funkcją Höldera dla funkcji f nazywa się funkcję, która każdemu punktowi $x \in D$ przyporządkowuje liczbę $\alpha_f(x)$.

Niech $H_t : [0,\infty) \to (0,1)$ będzie funkcją Höldera o wykładniku $\alpha > 0$. Multiułamkowym procesem ruchu Browna z parametrem funkcyjnym H_t jest nazywany proces stochastyczny $B_{H_t}(t)$ zdefiniowany dla $t \geq 0$ wzorem:

$$B_{Ht}(t) = \frac{1}{\Gamma\left(H_t + \frac{1}{2}\right)} \left\{ \int_{-\infty}^{0} \left[(t-s)^{H_t - 1/2} - (-s)^{H_t - 1/2} \right] dB(s) + \int_{0}^{t} (t-s)^{H_t - 1/2} \, dB(s) \right\} \quad (8)$$

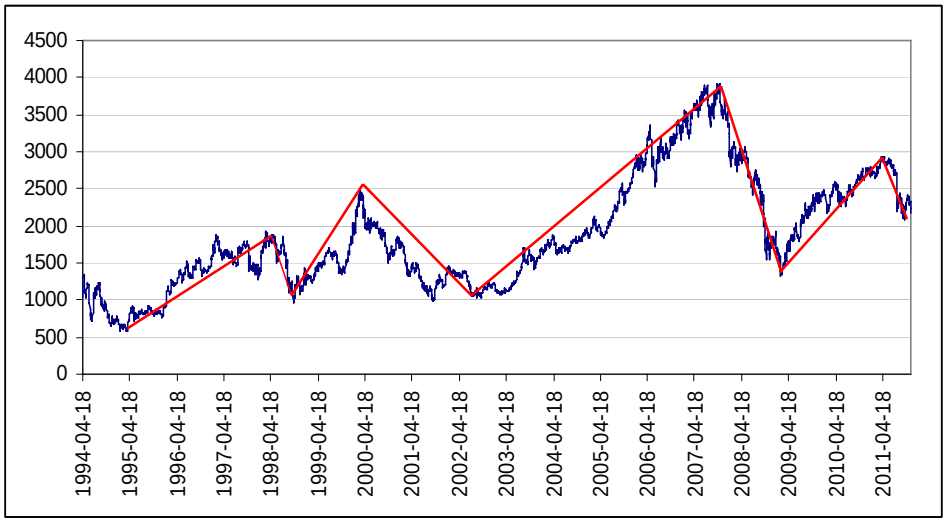
gdzie B jest standardowym procesem ruchu Browna.



Rys. 3. Multiułamkowy proces ruchu Browna dla funkcji Höldera w postaci $H(t) = (\sin^2(6t) + 0{,}1) / 1{,}3$

3. Podejmowanie decyzji z uwzględnieniem powyższych teorii

Do analizy empirycznej posłużą dane indeksu GPW w Warszawie – WIG20 za okres 16.04.1994-21.11.2011. Widać podobieństwa rysunków teoretycznych 1 i 2 oraz rysunku empirycznego 4. Wyraźnie w badanym okresie są zaznaczone fale wzrostowe 1, 3 oraz 5, a także fale zniżkowe A oraz C. Można także wyodrębnić fale korygujące: 2, 4 oraz B.

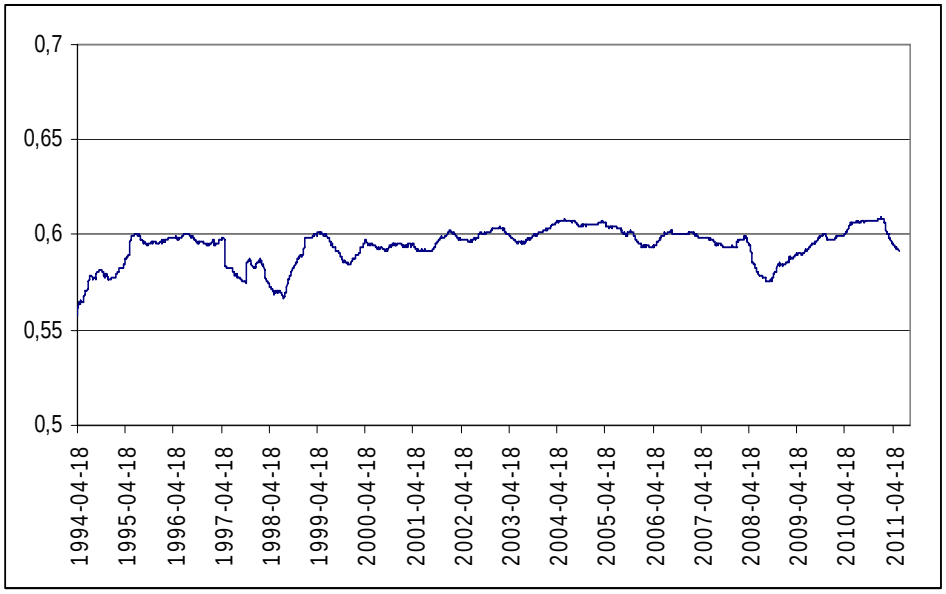


Rys. 4. Wartości kursu zamknięcia indeksu WIG20 w okresie 16.04.1994-21.11.2011

Źródło: Opracowanie własne na podstawie danych GPW w Warszawie.

Analizując długość fali A oraz korzystając z własności mówiącej, że długość fali C oblicza się mnożąc liczbę 0,618 i długość fali A, otrzymuje się informację, że fala zniżkowa C zakończy się około 4474 notowania (około lutego 2012 roku) na poziomie około 1372. Jeżeli natomiast ma miejsce płaska korekta, wówczas długość fali C stanowi 1,618 fali A, czyli fala C zakończy się około 4818 notowania (około lipca 1013 roku), miałaby ona jednak poziom ujemny! Podobnie można analizować fale krótsze oraz liczby z ciągu Fibonacciego.

Dla danych indeksu WIG20 obliczono metodą przeskalowanego zakresu wykładnik Hursta. Jego wartość dla całego szeregu to 0,594657, a zatem szereg jest persystentny, występuje w nim pamięć, jednakże nie jest on stacjonarny, co można zbadać za pomocą testu. Świadczy to także o braku stacjonarności, np. nieregularny kształt wyestymowanej funkcji Höldera. Punktowe wykładniki Höldera przedstawia rysunek 5. Wykładniki zawierają się w przedziale (0,56;0,61). Świadczy to o pamięci w szeregu, czyli na obserwacje przyszłe mają wspływ obserwacje historyczne. Z prawdopodobieństwem bliskim 0,6 po wzroście wartości indeksu nastąpi kolejny wzrost (po spadku kolejny spadek), z prawdopodobieństwem 1-0,6 trend się odwróci.



Rys. 5. Punktowe wykładniki Höldera dla danych WIG20 w okresie 16.04.1994-21.11.2011

Biorąc pod uwagę wielkości wyznaczone za pomocą fal Elliotta, jeszcze przez kilka miesięcy będzie wystepował trend spadkowy, patrząc zaś na wartości funkcji Höldera można dodać, że z niewielkimi wahaniami wartości indeksowych.

Podsumowanie

Decydent podejmujący działalność inwestycyjną kieruje się chęcią osiągnięcia maksymalnego zysku przy minimalnym ryzyku. Racjonalny inwestor może w tym działaniu posiłkować się różnymi metodami (w tym ilościowymi) umożliwiającymi analizy i wspomagającymi proces podejmowania decyzji.

Połączenie wybranych metod ilościowych jest uzasadnione, jednakże czas pokaże, na ile uwzględnienie powyższych metod będzie przydatne w trakcie podejmowania decyzji oraz zarządzania ryzykiem inwestycyjnym. Z pewnością jednak opisane powyżej metody, biorąc pod uwagę ich własności geometryczne, są interesujące dla matematyka i przydatne dla inwestora.

Literatura

1. Benassi A., Jaffard S., Roux D.: *Gaussian Processes and Pseudodifferential Elliptic Operators*. „Rev. Mat. Iberoamericana" 1997, 13 (1), s. 19-89.
2. Daoudi K., Lévy Véhel J., Meyer Y.: *Construction of Continuous Functions with Prescribed Local Regularity*. „Journal of Constructive Approximations" 1998, 014(03), s. 349-385.

3. Falconer K.J.: *Fractal Geometry, Mathematical Foundations and Applications*. John Wiley & Sons, New York 1990.
4. Frost A.J., Prechter R.R: *Teoria fal Elliotta*. Wydawnictwo WIG-PRESS, Warszawa 1995.
5. Jaworski J.: *Teoria i praktyka zarządzania finansami przedsiębiorstw*. CeDeWu, Warszawa 2010.
6. Lewandowski A., Michalski T.: *Analiza techniczna rynku kapitałowego*. Szkoła Główna Handlowa, Warszawa 2001.
7. Mandelbrot B.B.: *Fractals and Scaling in Finance. Discontinuity, Concentration, Risk*. Springer-Verlag, New York 1997.
8. Mandelbrot B.B., Van Ness J.W.: *Fractional Brownian Motion, Fractional Noises and Applications*. „SIAM Review" 1968, Vol. 10, No. 4, s. 422-437.
9. Nowakowski J., Borowski K.: *Zastosowanie teorii Carolana i Fischera na rynku kapitałowym*. DIFIN, Warszawa 2005.
10. Peltier R.F., Lévy Véhel J.: *Multifractional Brownian Motion: Definition and Preliminary Results*. INRIA Recquencourt, Rapport de recherche No. 2645, 1995.
11. Surdel P.: *Forex – analiza techniczna*. Złote Myśli, Warszawa 2004.

ELLIOTT THEORY AND FRACTALS – SIMILARITIES AND DIFFERENCES BETWEEN MODELLING FINANCIAL TIME SERIES AND DESCRIPTIONS OF INVESTORS DECISION MAKING

Summary

The article presents theoretical basis and practical applications of selected quantity methods that can be used in modeling financial time series, where elements of Elliott theory and fractal geometry are included. The aim of this work is to present models to support the investor in decision making, which includes new market tendencies. The process of investing into financial markets is a dynamic process depending on frequent changes, that direction and impact is difficult to predict in the long periods of time. This work shows theoretical basis of used methods and results of carried out empirical analyses.

Monika Miśkiewicz-Nawrocka

MODELE EKONOMICZNE Z DYNAMIKĄ CHAOTYCZNĄ

Wprowadzenie

Od czasu pojawienia się w literaturze pojęcia deterministycznego chaosu można znaleźć wiele przykładów układów dynamicznych (zarówno z czasem ciągłym, jak i z czasem dyskretnym) o chaotycznej dynamice. Są wśród nich układy z różnych dziedzin nauki: matematyczne, fizyczne, chemiczne, biologiczne, medyczne, a także ekonomiczne.

Do opisu zjawisk ekonomicznych od połowy XX wieku stosuje się nieliniowe modele deterministyczne. Podejście takie zaproponowali M. Kalecki, J. Tinberger i N. Kaldor, którzy do opisu cykli koniunkturalnych [7, s. 261-273] wykorzystali modele deterministyczne. Jednak ich modele nie opisywały zbyt dobrze złożonej dynamiki zjawisk ekonomicznych [17, s. 113]. W 1975 roku R. May i J.R. Beddington [13, s. 35, za: 18] zasygnalizowali możliwość zastosowania teorii chaosu w ekonomii. Od tego czasu zbudowano wiele nowych modeli ekonomicznych z dynamiką chaotyczną oraz zidentyfikowano chaos w wielu już istniejących. Nieliniowe układy chaotyczne wzbudziły tak duże zainteresowanie wśród ekonomistów, ponieważ pozwalają generować szeregi o bardzo skomplikowanej dynamice.

W artykule zostaną zaprezentowane wybrane modele ekonomiczne o chaotycznej dynamice.

1. Chaos deterministyczny

Początki teorii chaosu sięgają końca XIX wieku i są związane z pracami francuskiego matematyka Henri Poincaré. Poincaré, badając zachowanie się pojedynczych trajektorii trzech ciał niebieskich, odkrył istnienie bardzo złożonych struktur – trajektorii chaotycznych [19, s. 11]. Do prekursorów teorii deterministycznego chaosu można również zaliczyć: P. Fatou, G.M. Julia (początek XX wieku) [18, s. 16], G. Birkhoffa (lata dwudzieste XX wieku), M.L. Cartwighta i J.E. Littlewooda (lata czterdzieste), S. Smale'a, A.N. Kołmogorowa i jego

współpracowników (lata sześćdziesiąte) [19, s. 11]. Prowadzone przez nich badania wykazały nieznane wcześniej własności dynamiki nieliniowej oraz bardzo skomplikowaną i nieregularną dynamikę prostych nieliniowych układów dynamicznych [18, s. 16]. Wśród wielu badaczy różnych dyscyplin naukowych teorię chaosu najbardziej rozpowszechniło odkrycie Edwarda Lorenza w latach siedemdziesiątych*. Lorenz odkrył podstawową cechę nieliniowych układów chaotycznych – wrażliwość układu na zmianę warunków początkowych. Rozgłos teorii chaosu i jej zastosowaniom do modelowania układów dynamicznych dodatkowo nadała praca M.J. Feigenbauma, który jako pierwszy wykorzystał komputery do badania słabo rozumianej w latach 70. nieliniowej dynamiki [15, s. 32, 286].

Termin „chaos deterministyczny" został wprowadzony w 1975 roku przez T.Y. Li i J.A. Yorke'a, jednak w literaturze można znaleźć wiele definicji chaosu w deterministycznych układach dynamicznych. Definicje te nie zawsze są równoważne (jednoznaczne), ponieważ wywodzą się z różnych dyscyplin matematycznych, tj. teoria równań różniczkowych i różnicowych, jakościowa teoria układów dynamicznych czy teoria ergodyczna [23, s. 82]. Pomimo wielości definicji deterministycznego chaosu spotykanych w literaturze, badacze (naukowcy) są zgodni, że prawidłowa definicja chaosu powinna dobrze oddawać naturę dynamiki chaotycznej, czyli zakładać istnienie dynamiki nieokresowej w badanym układzie deterministycznym, wrażliwość na zmianę warunków początkowych oraz istnienie pewnego istotnego mechanizmu deterministycznego odpowiedzialnego za rekurencyjne zachowanie się układu [16, s. 34].

Powszechnie stosowana definicja chaosu odwołuje się do twierdzenia T.Y. Li i J.A. Yorke'a o chaosie [23, s. 86]. Twierdzenie podaje stosunkowo łatwy do sprawdzenia warunek wystarczający istnienia dynamiki chaotycznej w układach dynamicznych o znanej funkcji generującej *f*. Najbardziej znanymi przykładami odwzorowań chaotycznych w sensie Li, Yorke'a są odwzorowanie logistyczne i odwzorowanie trójkątne. Niektóre z powstałych modeli ekonomicznych zostały skonstruowane tak, aby spełniały założenia i warunki twierdzenia Li, Yorke'a [23, s. 188-189]. Przykładem jest neoklasyczny model wzrostu Daya [5, s. 406--414]. Ponadto powstały również modele generowane przez chaotyczne odwzorowania matematyczne, w szczególności przez funkcję logistyczną. Przykładami są model wzrostu Havelmo-Stutzera [22, s. 253-276; 23, s. 192-197; 9, s. 122] oraz chaotyczny model popytu konsumpcyjnego Benhabiba i Daya [1, s. 113--118; 2, s. 37-55; 23, s. 199-201].

Podstawowym atrybutem dynamiki chaotycznej jest wrażliwość układu na zmianę warunków początkowych. Definicja odwołująca się do pojęcia wrażli-

* Lorenz rozpoczął swoje prace w 1960 roku, ale społeczność naukowa doceniła je dopiero w latach 70. ubiegłego wieku.

wości została sformułowana przez R.L. Devaneya [6, za: 23, s. 88] w 1987 roku. Według tej definicji układ dynamiczny (X, f) jest układem chaotycznym w zbiorze *X*, jeśli odwzorowanie *f* jest wrażliwe na zmianę warunków początkowych i topologicznie tranzytywne oraz zbiór punktów okresowych odwzorowania *f* jest gęsty w zbiorze *X*.

Miarą wspomnianej wrażliwości układu na zmianę warunków początkowych są wykładniki Lapunowa, które mierzą średnie tempo rozchodzenia się w przestrzeni stanów trajektorii początkowo bliskich sobie punktów. Im większa wartość dodatniego największego wykładnika Lapunowa, tym większa wrażliwość układu na zmianę warunków początkowych, a także większy poziom chaosu. Istnienie w układzie więcej niż jednego dodatniego wykładnika określa się mianem hiperchaosu [12, s. 241]. Niektórzy autorzy uznają istnienie dodatniego największego wykładnika Lapunowa za warunek konieczny i wystarczający istnienia chaosu w układzie [9, s. 103-133].

Trajektorie typowego układu chaotycznego tworzą w przestrzeni stanów bardzo złożoną strukturę, tzw. dziwny atraktor, która nie wypełnia jej w sposób przypadkowy. Pojęcie „dziwnego atraktora" również jest stosowane do definiowania chaosu – układ jest chaotyczny, gdy ma dziwny atraktor [10, s. 88].

W praktyce do identyfikacji chaosu w procesach rzeczywistych wielu autorów posługuje się definicją mówiącą, że układ dynamiczny jest chaotyczny, gdy jest wrażliwy na zmianę warunków początkowych [23, s. 161; 20, za: 18, s. 19]. Zaletą tej definicji jest możliwość zweryfikowania jej za pomocą największego wykładnika Lapunowa oraz wymiaru korelacyjnego. Natomiast jej wadą jest fakt, że odnosi się tylko do układów dyssypatywnych. W przypadku układów konserwatywnych do badania regularności dynamiki wykorzystuje się pojęcie entropii Kołomogorowa. Według tej definicji układ dynamiczny jest chaotyczny, gdy ma dodatnią, skończoną entropię [18, s. 19]. Nieskończona entropia oznacza, że układ jest losowy, natomiast entropia równa 0 oznacza, że układ jest deterministyczny.

2. Przykłady chaotycznych układów dynamicznych

2.1. Model Kaldora [7, s. 264-265; 18, s. 38-40]

Jednym z najwcześniejszych nieliniowych modeli ekonomicznych jest opublikowany przez N. Kaldora [11, s. 78-92] w 1940 roku model cyklu koniunkturalnego. Kaldor rozwinął i zmodyfikował model zaprezentowany przez M. Kaleckiego w pracy *Próba teorii koniunktury* w 1933 roku[*]. Model Kaldora układu dynamicznego z czasem ciągłym ma postać:

[*] Kalecki przedstawił swoje dzieło języku niemieckim na III Europejskiej Konferencji Towarzystwa Ekonometrycznego w Leyden w 1933 roku, a w 1935 roku *Próba teorii koniunktury* została opublikowana w języku angielskim i francuskim [7, s. 262].

$$\dot{Y} = \alpha\big(I(Y,K) - S(Y,K)\big), \tag{1}$$

$$\dot{K} = I(Y,K) - \delta K\,, \tag{2}$$

gdzie:
$Y = Y(t)$ – wielkość produkcji,
$K = K(t)$ – wartość kapitału zaangażowanego w produkcję,
$I = I(t)$ – poziom inwestycji,
$S = S(t)$ – poziom oszczędności,
$\delta > 0$ – współczynnik spadku kapitału,
α – parametr.

W swoim modelu Kaldor przyjął następujące założenia:

$$I_Y > 0\,,\ I_K < 0\,,\ 0 < S_Y < 1\,,\ S_K > 0\,, \tag{3}$$

oraz istnieje taki poziom produkcji $\overline{Y}$, że:

$$I_{YY} > 0 \text{ dla } Y < \overline{Y}\,, \tag{4}$$

oraz:

$$I_{YY} < 0 \text{ dla } Y > \overline{Y}\,, \tag{5}$$

gdzie dolne indeksy oznaczają pochodne odpowiadające *i*-temu argumentowi. Przedstawiony model generuje zachowania cykliczne na podstawie zmiennych endogenicznych.

Przyjmując $\dot{Y} = \Delta Y_{t+1}$, $\dot{K} = \Delta K_{t+1}$, otrzymujemy dyskretną wersję modelu Kaldora:

$$\Delta Y_{t+1} = \alpha\big(I_t(Y_t,K_t) - S_t(Y_t)\big), \tag{6}$$

$$\Delta K_{t+1} = I_t(Y_t,K_t) - \delta K_t\,. \tag{7}$$

Ponieważ $\Delta Y_{t+1} = Y_{t+1} - Y_t$, $\Delta K_{t+1} = K_{t+1} - K_t$, więc:

$$Y_{t+1} = \alpha\big(I_t(Y_t,K_t) - S_t(Y_t)\big) + Y_t\,, \tag{8}$$

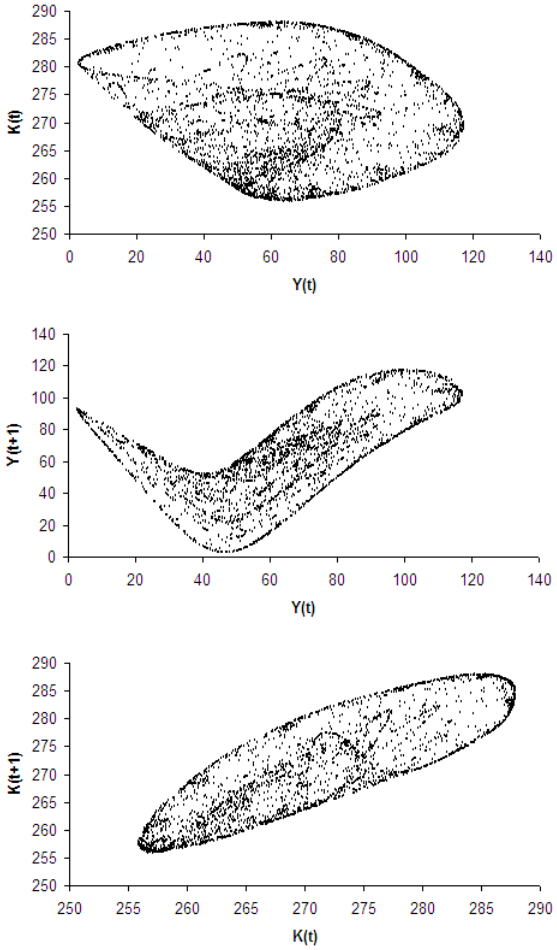
$$K_{t+1} = I_t(Y_t,K_t) + (1-\delta)K_t\,. \tag{9}$$

W powyższym modelu funkcja oszczędności I_t może mieć postać:

$$I_t = c \cdot 2^{\frac{-1}{(dY_t+\varepsilon)^2}} + eY_t + a\left(\frac{f}{K_t}\right)^g, \tag{10}$$

gdzie $a, c, d, e, f, g, \varepsilon$ są parametrami. Funkcja (10) spełnia założenia Kaldora (3). Dla małych wartości parametru *a* trajektorie układu Kaldora zmierzają do punktu stałego, dla większych wartości *a* do cyklu, natomiast dla *a* przekraczającego pewną wartość krytyczną do dziwnego atraktora [18, s. 39].

Na rysunku 1 przedstawiono dziwny atraktor modelu Kaldora dla wartości parametrów: $a = 0{,}2$, $c = 20$, $d = 0{,}01$, $e = 0{,}05$, $f = 280$, $g = 4{,}5$, $s = 0{,}21$, $\alpha = 20$, $\delta = 0{,}05$, $\varepsilon = 0{,}00001$ oraz dla stanu początkowego $(Y_0, K_0) = (65{,}265)$ w przestrzeniach (Y_t, K_t), (Y_t, Y_{t+1}), (K_t, K_{t+1}).



Rys. 1. Dziwny atraktor w modelu Kaldora w przestrzeniach (Y_t, K_t), (Y_t, Y_{t+1}), (K_t, K_{t+1})

2.2. Model CARAL [3, s. 145-170, za: 4, s. 53-56]

Model CARAL (Constant Absolute Risk Aversion z funkcją produkcji Leontiewa) należy do szerokiej klasy modeli nakładających się pokoleń OLG.

Rozważmy gospodarkę złożoną z dwóch pokoleń: ludzi młodych i ludzi starszych. Członkowie każdego pokolenia żyją i konsumują w dwóch okresach swojego życia (młodość i starość), ale pracują tylko w pierwszym z nich, kiedy są młodzi. Niech c_t oraz c_{t+1} oznaczają poziomy konsumpcji w obu okresach, natomiast l_t – poziom pracy w okresie *t*. Głównym problemem młodego pokolenia na początku okresu *t* jest wybranie poziomów konsumpcji c_t i c_{t+1} oraz poziomu pracy l_t, tak aby maksymalizować ogólną użyteczność przy pewnych ograniczeniach budżetowych.

Oznaczmy przez $u_1(c_t)$ oraz $u_2(c_{t+1})$ funkcje użyteczności dla poziomu konsumpcji odpowiednio w pierwszym i drugim okresie. Niech $v(l_t)$ oznacza użyteczność pracy (poziom niezadowolenia z każdej wykonanej jednostki pracy). Załóżmy, że powyższe funkcje użyteczności wyrażają się następującymi wzorami:

$$u_1(c_t) = -re^{-c_t},\ r > 0, \tag{11}$$

$$u_2(c_{t+1}) = \frac{1}{\alpha} c_t^{\alpha},\ 0 < \alpha < 1, \tag{12}$$

$$v(l_t) = \frac{1}{\gamma} l_t^{\gamma},\ \gamma > 1. \tag{13}$$

Pierwsze wyrażenie opisuje funkcję użyteczności o stałej bezwzględnej awersji do ryzyka, natomiast dwa pozostałe o stałej względnej awersji do ryzyka. Rozwiązując wyżej postawiony problem optymalizacyjny, dynamika poziomu konsumpcji wyraża się wzorem[*]:

$$c_{t+1} = \left(l_t^{\gamma} - rc_t e^{-c_t}\right)^{1/\alpha}. \tag{14}$$

Powyższe równanie opisuje optymalną ewolucję poziomu konsumpcji pochodzącą z międzyokresowych wyborów konsumenta pomiędzy konsumpcją a czasem wolnym.

Wprowadzając do modelu liniową funkcję produkcji Leontiewa, dynamika pracy jest dana wzorem:

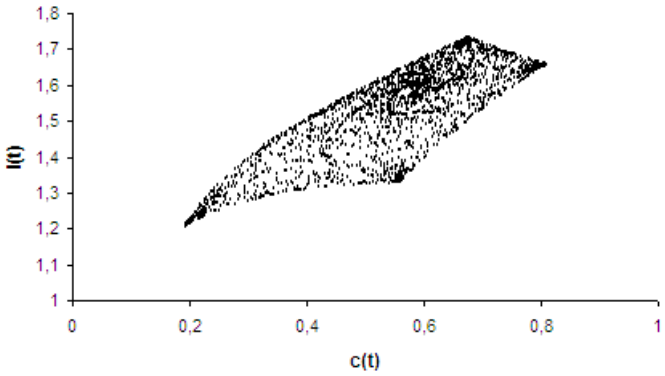
[*] Szczegółowe rozwiązanie zob. [14].

$$l_{t+1} = b(l_t - c_t). \tag{15}$$

Powyższe równania opisują ewolucję układu dynamicznego, gdzie parametr $\gamma > 1$ oznacza elastyczność użyteczności pracy, $0 < \alpha < 1$ jest elastycznością użyteczności przyszłej konsumpcji, b jest współczynnikiem produktywności, natomiast r oznacza „stromość" wykładniczej funkcji użyteczności.

Medio i Negroni [14] pokazali, że układ CARAL (14)-(15) ma punkt równowagi, który może zostać zaburzony poprzez zmianę wartości parametrów modelu. W szczególności, przy odpowiedniej zmianie parametrów, zachowanie dynamiki modelu pokazuje bardzo bogaty scenariusz, pojawiają się zachowania okresowe, nieokresowe, a nawet chaotyczne.

Dla wartości parametrów $\gamma = 1{,}2$, $\alpha = 0{,}49$, $r = 3$ oraz $b = 1{,}57$ dynamika układu (14)-(15) jest chaotyczna. Przy $t \to \infty$ trajektorie punktów zmierzają do ograniczonego zbioru, tworząc chaotyczny atraktor. Chaotyczny atraktor układu CARAL dla powyższych zaprezentowano na rysunku 2.



Rys. 2. Chaotyczny atraktor układu CARAL (14)-(15) dla stanu początkowego $c_0 = 0{,}3,\ l_0 = 1{,}3$

2.3. Model rynku pracy [8, s. 470-483]

Rozważmy pewną gospodarkę z wyróżnionym przedsiębiorstwem oraz wyróżnionym robotnikiem-konsumentem. Funkcja produkcji przedsiębiorstwa jest funkcją Cobba-Douglasa. Zakładając, że kapitał ma stałą wartość unormowaną do 1, funkcja produkcji wyraża się wzorem:

$$Y = DL^a,\ 0 < a < 1,\ D > 0, \tag{16}$$

gdzie D jest parametrem odzwierciedlającym stały postęp technologiczny.

Niech π oraz ω oznaczają odpowiednio całkowity zysk przedsiębiorstwa oraz stawkę płac. Załóżmy, że cena jednostki wytworzonego towaru *p* jest równa jeden. Wówczas funkcja zysku ma postać:

$$\pi = DL^a - \omega L \,. \tag{17}$$

Rozwiązując zadanie maksymalizacji zysku (4.19), otrzymujemy funkcję popytu na pracę daną wzorem:

$$L^D = f_1(\omega) = \left(\frac{\omega}{aD}\right)^{\frac{1}{a-1}}. \tag{18}$$

Robotnik-konsument maksymalizuje swoją użyteczność wyrażoną za pomocą funkcji CES:

$$u(C,L) = \left[C^b + (N-L)^b\right]^{\frac{1}{b}},\ b \in (-\infty,1), \tag{19}$$

gdzie $N > 0$ jest maksymalną wielkością podaży siły roboczej, a $C > 0$ – poziomem konsumpcji konsumenta.

Załóżmy, że $N = 1$. Rozwiązując zadanie maksymalizacji funkcji użyteczności (19), przy ograniczeniu budżetowym $C = \omega L$, otrzymujemy optymalną funkcję podaży siły roboczej:

$$L^S = f_2(\omega) = \frac{1}{1 + \omega^{\frac{b}{b-1}}}. \tag{20}$$

Niech $z(\omega) = L - S$ oznacza nadwyżkowy popyt na pracę, gdzie *L* i *S* są bieżącą wartością popytu i podaży pracy. Płaca jest uporządkowana w sposób ciągły przez bieżący nadwyżkowy popyt na pracę zgodnie z następującą regułą:

$$\dot{\omega} = lz(\omega) = l(L - S),\ l > 0\,, \tag{21}$$

gdzie *l* jest stałą.

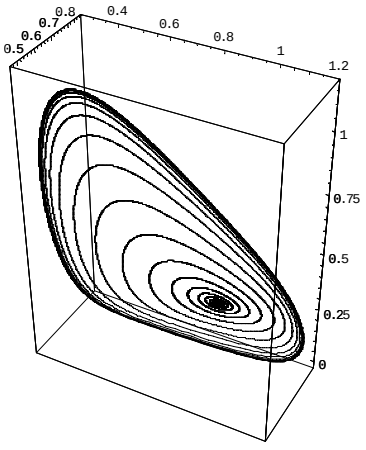
Ostatecznie model ma postać:

$$\dot{L} = g\left(L^D - L\right) = g\left(\left(\frac{\omega}{aD}\right)^{\frac{1}{a-1}} - L\right) = g\left(f_1(\omega) - L\right),\ g > 0\,, \tag{22}$$

$$\dot{S} = d\left(L^S - S\right) = d\left(\frac{1}{1+\omega^{\frac{b}{b-1}}} - S\right) = d\left(f_2(\omega) - S\right),\ d > 0\,, \tag{23}$$

$$\dot{\omega} = l\left(L - S\right),\ l > 0\,. \tag{24}$$

L. Fanti i P. Manfredi przeprowadzili numeryczne symulacje pokazujące, jak zmieniają się portrety fazowe powyższego modelu dla różnych wartości parametru *g*, przyjmując $D = 1$, $a = 0{,}15$, $b = -10$, $d = 4$, $l = 4$. Symulacje były przeprowadzone w otoczeniu punktu równowagi $E = \left(L^*, S^*, \omega^*\right) = \left(0.832, 0.832, 0.182\right)$. Warunkiem początkowym był punkt $E_1 = \left(0.825, 0.835, 0.18\right)$. Okazało się, że dla wartości $g > 1.068$ punkt E_1 jest lokalnie asymptotycznie stabilny, w szczególności dla $1.068 < g < 1.3$ E_1 jest stabilnym ogniskiem, a następnie staje się stabilnym węzłem. Dla $0{,}84 < g < 1.068$ trajektorie punktów położonych wystarczająco blisko punktu E_1 początkowo oddalają się od siebie, a następnie są zbieżne do stabilnego cyklu granicznego. W przypadku gdy $0.82 < g < 0.84$, cykl wykazuje małe oscylacje dla popytu i podaży. Natomiast gdy $0.62 < g < 0.82$, trajektorie punktów „wędrują" w sposób nieregularny, losowy w ograniczonym obszarze na płaszczyźnie $\left(L, S\right)$. Przy $t \to \infty$ trajektorie punktów zmierzają do ograniczonego zbioru, tworząc chaotyczny atraktor. Dla $0.53 < g < 0.62$ autorzy otrzymali quasi-okresowy atraktor, a dla $g < 0{,}53$ globalną niestabilność.



Rys. 3. Chaotyczne trajektorie układu (22)-(24) na płaszczyźnie (S, L) dla *g* = 0,8. Warunek początkowy $L^0 = 0{,}825$, $S^0 = 0{,}835$, $\omega_0 = 0{,}18$

Źródło: Opracowanie własne na podstawie: [8, s. 476].

Podsumowanie

Pomimo iż pojęcie deterministycznego chaosu pojawiło się w literaturze blisko 30 lat temu, ciągle brak jest jednej jednoznacznej definicji. Najbardziej odpowiednią, oddającą naturę dynamiki chaotycznej, wydaje się być definicja podana na międzynarodowej konferencji na temat chaosu, zorganizowanej przez Royal Statistical Society w 1986 roku, gdzie chaos zdefiniowano jako „stochastyczne zachowanie występujące w układzie deterministycznym" [21, za: 18, s. 16]. Jednak definicja ta jest nieformalna. Zasygnalizowana przez R. Maya i J.R. Beddingtona możliwość zastosowania teorii chaosu w ekonomii spowodowała powstanie wielu nowych modeli ekonomicznych z dynamiką chaotyczną lub odkrycie chaosu w już istniejących modelach ekonomicznych. W artykule dokonano przeglądu definicji deterministycznego chaosu oraz zademonstrowano chaotyczną dynamikę wybranych ekonomicznych modeli dynamicznych. Zdaniem autorki warto byłoby podjąć badania nad ulepszeniem i uzupełnieniem listy modeli procesów ekonomicznych z dynamiką chaotyczną.

Literatura

1. Benhabib J., Day R.: *Erratic Accumulation*. „Economics Letters" 1980, Vol. 6, s. 113-118.
2. Benhabib J., Day R.: *A Characterization of Erratic Dynamics in the Overlapping Generation Model*. „Journal of Economic Dynamics and Control" 1982, Vol. 4, s. 37-55.
3. Benhabib J., Laroque G.: *On Competitive Cycles in Productive Economy*. „Journal of Economic Theory" 1988, Vol. 45, s. 145-170.
4. Bordignon S., Lisi F.: *Predictive Accuracy for Chaotic Economic Models*. „Economics Letters" 2001, 70, s. 51-58.
5. Day R.: *Irregular Growth Cycles*. „American Economic Review" 1982, Vol. 72, s. 406-414.
6. Devaney R.L.: *An Introduction to Chaotic Dynamical Systems*. Addison-Wesley Publishing Company, Inc., Redwood City 1987.
7. Drabik E.: *Dynamiczne nieliniowe modele ekonometryczne: model cykli koniunkturalnych Kaleckiego-Kaldora oraz model wzrostu*. W: *Rynek kapitałowy. Skuteczne inwestowanie*. Red. W. Tarczyński. Szczecin 2002, s. 261-273.
8. Fanti L., Manfredi P.: *Neoclassical Labour Market Dynamics, Chaos and the Real Wage Philips Curve*. „Journal of Economic Behavior & Organization" 2007, Vol. 62, s. 470-483.
9. Frank M., Stengos T.:*Chaotic Dynamics in Economic Time-Series*. „Journal of Economic Surveys" 1988, Vol. 2, s. 103-133.
10. Garrido L.: *Dynamical Systems of Chaos*. „Lectures Notes in Physics" 1983, Vol. 179.

11. Kaldor N.: *A Model of the Trade Cycle*. „Economic Journal" 1940, Vol. 50, s. 78-92.
12. Kantz H., Schreiber T.: *Nonlinear Time Series Analysis*. Cambridge University Press, Cambridge 2004 (second edition).
13. May R., Beddington J.R.: *Nonlinear Difference Equations: Stable Points, Stable Cycles, Chaos*. Maszynopis, 1975.
14. Medio A., Negroni G.: *Chaotic Dynamics in Overlapping Models with Production*. W: *Nonlinear Dynamics and Economics*. Red. W.A. Barnett, A.P. Kirman, M. Salmon. Cambridge University Press, Cambridge 1996.
15. Morrison F.: *Sztuka modelowania układów dynamicznych: deterministycznych, chaotycznych, stochastycznych*. Wydawnictwa Naukowo-Techniczne, Warszawa 1996.
16. Nowiński M.: *Nieliniowa dynamika szeregów czasowych*. Wydawnictwo Akademii Ekonomicznej, Wrocław 2007.
17. Orzeszko W.: *Metody identyfikacji i prognozowania chaotycznych szeregów czasowych*. W: *Metody ilościowe w naukach ekonomicznych*. Czwarte Warsztaty Doktorskie z Zakresu Ekonometrii i Statystyki. Red. A. Welfe. Szkoła Główna Handlowa, Warszawa 2004, s. 113-135.
18. Orzeszko W.: *Identyfikacja i prognozowanie chaosu deterministycznego w ekonomicznych szeregach czasowych*. Polskie Towarzystwo Ekonomiczne, Warszawa 2005.
19. Ott E.: *Chaos w układach dynamicznych*. Wydawnictwa Naukowo-Techniczne, Warszawa 1997.
20. Sosvilla-Rivero S., Fernandez-Rodrigez F., Andrada-Felix J.: *Testing Chaotic Dynamics via Lyapunov Exponents*. Working Papers 2000-07, FEDEA.
21. Stewart I.: *Czy Bóg gra w kości? Nowa matematyka chaosu*. Wydawnictwo Naukowe PWN, Warszawa 1994.
22. Stutzer M.T.: *Chaotic Dynamics and Bifurcation in a Macro-model*. „Journal of Economic Dynamics and Control" 1980, Vol. 2, s. 253-276.
23. Zawadzki H.: *Chaotyczne systemy dynamiczne. Elementy teorii i wybrane zagadnienia ekonomiczne*. Zeszyty Naukowe Akademii Ekonomicznej, Katowice 1996.

THE ECONOMIC MODELS WITH CHAOTIC DYNAMICS

Summary

Since 1975, when the R. May and J.R. Beddington informed about the possibility of application of chaos theory in economics, built many new economic models with chaotic dynamics and chaos have been identified in a number of already existing models.

This paper presents briefly the theory of deterministic chaos and properties of chaotic dynamic. In addition, presents some economic models of chaotic dynamics.

Ewa Pośpiech

PODPISY NIEZAPRZECZALNE – PODPISY Z DODATKOWĄ FUNKCJONALNOŚCIĄ

Wprowadzenie

Powszechna informatyzacja powoduje, że w coraz większym stopniu są dostępne różne możliwości przeprowadzania operacji w sieci. Standardem jest komunikacja elektroniczna, gromadzenie i zarządzanie danymi, zakupy internetowe, transakcje w systemie bankowości elektronicznej, a coraz więcej użytkowników korzysta m.in. z różnego rodzaju usług e-urzędów. Jednym z najistotniejszych aspektów związanym z tego typu operacjami jest bezpieczeństwo; w kwestii tej znaczący udział ma współczesna kryptografia, która wykorzystuje narzędzia matematyczne, a której sprzyja dynamicznie rozwijająca się technika komputerowa.

W przypadku różnych transakcji zachodzi konieczność potwierdzenia ich wykonania, co jest realizowane przez składanie podpisu. W przypadku operacji wirtualnych jest składany podpis cyfrowy. Najpowszechniejszym zastosowaniem podpisów cyfrowych jest podpisywanie wiadomości przesyłanych drogą elektroniczną. Tą drogą odbywa się wymiana dokumentów z różnymi instytucjami i partnerami biznesowymi, składa się podania administracyjne, uwierzytelnia e-faktury, dokonuje się uwierzytelniania transakcji w systemie bankowości elektronicznej itp. Innym, równie istotnym zastosowaniem podpisów cyfrowych jest ich składanie w celu zabezpieczenia oprogramowania.

Celem składania podpisu cyfrowego jest przede wszystkim zagwarantowanie autentyczności, niezaprzeczalności i integralności wiadomości (dokumentu). Inną cechą, która w pewnych sytuacjach może być wymagana, jest dodatkowa funkcjonalność podpisu. Najczęściej oznacza to połączenie algorytmu podpisu cyfrowego z charakterystycznym protokołem w celu uzyskania dodatkowych własności, których zwykły algorytm nie zapewnia. Przykładem takich algorytmów są schematy podpisu niezaprzeczalnego. Schematy te, w odróżnieniu od zwykłych schematów podpisów cyfrowych, wymagają w procesie weryfikacji podpisu współpracy z podmiotem podpisującym.

Celem tego artykułu jest przedstawienie wybranych protokołów kryptograficznych umożliwiających składanie podpisów niezaprzeczalnych, które znajdują zastosowania m.in. w zabezpieczaniu oprogramowania przed kopiowaniem i rozpowszechnianiem wśród osób nieuprawnionych.

1. Podstawy teoretyczne

Bazą konstrukcji algorytmów kryptograficznych są pojęcia matematyczne. Głównie wykorzystuje się zagadnienia związane z teorią liczb oraz algebrą abstrakcyjną. Dlatego też przytoczono wybrane zagadnienia teoretyczne wykorzystywane przy formułowaniu prezentowanych protokołów podpisu niezaprzeczalnego (na podstawie [1; 3; 5]).

Definicja 1. Niech *n* jest liczbą całkowitą dodatnią oraz *a, b* są liczbami całkowitymi. Mówimy, że *a* przystaje do *b* modulo *n*, jeśli liczba *a* i liczba *b* dają takie same reszty przy dzieleniu przez *n*. Zależność tę zapisujemy w postaci wyrażenia zwanego kongruencją:

$$a \equiv b \pmod{n}.$$

Definicja 2. Niech *n* będzie liczbą naturalną oraz *a* liczbą całkowitą. Element $x \in \mathbf{Z}$ ($\mathbf{Z}$ – zbiór liczb całkowitych) nazywa się elementem odwrotnym do *a* modulo *n*, jeśli jest spełniona kongruencja:

$$ax \equiv 1 \pmod{n}.$$

Definicja 3. Grupą multiplikatywną grupy $\mathbf{Z}_n$, gdzie $\mathbf{Z}_n$ jest zbiorem w postaci $\mathbf{Z}_n = \{0, 1, 2, ..., n-1\}$, jest $\mathbf{Z}_n^* = \{a \in \mathbf{Z}_n : \mathrm{NWD}(a, n) = 1\}$. W szczególności gdy *n* jest liczbą pierwszą, wówczas $\mathbf{Z}_n^* = \{a \in \mathbf{Z} : 1 \le a \le n-1\}\}$.

Definicja 4. Rzędem grupy *G* lub podgrupy *P* nazywamy liczbę jej elementów.

Definicja 5. Jeśli $G = \{g^k : k \in \mathbf{Z}\}$ dla pewnego $g \in G$, gdzie *G* jest grupą, to *G* nazywamy grupą cykliczną, natomiast element *g* nazywamy generatorem grupy *G*.

Definicja 6. Niech G jest grupą oraz $g \in G$. Jeśli istnieje liczba całkowita dodatnia r taka, że zachodzi $g^r = 1$, to najmniejszą taką liczbę nazywa się rzędem elementu g w grupie G. Oznaczamy $r = \text{rząd}_G a$.

Definicja 7. Niech n będzie liczbą naturalną oraz a liczbą całkowitą taką, że NWD(a, n) = 1. Rzędem liczby a modulo n nazywamy najmniejszą liczbę naturalną r taką, że $a^r \equiv 1 \pmod{n}$.

Twierdzenie 1. Niech $g \in G$ (G – grupa multiplikatywna) i niech k, l będą liczbami całkowitymi. Wówczas $g^l = g^k$ wtedy i tylko wtedy, gdy $l \equiv k \pmod{\text{rząd}_G g}$.

Definicja 8. Niech n będzie liczbą naturalną. Funkcją Eulera $\varphi(n)$ nazywamy liczbę elementów naturalnych k takich, że $k \leq n$ oraz względnie pierwszych z n.

Definicja 9. Niech $g \in \mathbf{Z}_n^*$. Jeśli rząd g wynosi $\varphi(n)$, to o g mówimy, że jest generatorem lub elementem (pierwiastkiem) pierwotnym $\mathbf{Z}_n^*$. Jeśli $\mathbf{Z}_n^*$ ma generator, to o $\mathbf{Z}_n^*$ mówi się, że jest cykliczna.

Definicja 10. Niech n będzie liczbą naturalną oraz a liczbą całkowitą taką, że NWD(a, n) = 1. Jeśli rząd elementu a modulo n wynosi $\varphi(n)$, to a nazywamy pierwiastkiem pierwotnym modulo n.

Definicja 11. Niech p będzie liczbą pierwszą oraz niech g będzie pierwiastkiem pierwotnym modulo p. Dla każdej liczby całkowitej $A \in \{1, 2, \ldots, p-1\}$ istnieje wykładnik $a \in \{0, 1, \ldots, p-2\}$ taki, że:

$$A \equiv g^a \pmod{p}.$$

Wykładnik ten nazywamy logarytmem dyskretnym liczby A przy podstawie g ($a = \log_g A$).

2. Podpisy niezaprzeczalne

Podpisy niezaprzeczalne są rodzajem podpisu, którego weryfikacja wymaga współpracy z osobą podpisującą – składający podpis ma kontrolę nad sprawdzaniem podpisu oraz nad podmiotami sprawdzającymi (osoby trzecie nie mogą uzyskać potwierdzenia autentyczności informacji/dokumentu bez konfrontacji z osobą podpisującą). Takie określenie podpisów niezaprzeczalnych nie tłuma-

czy ich „niezaprzeczalności"; nazwa wywodzi się stąd, że jeśli podmiot podpisujący byłby zmuszony do potwierdzenia albo wyparcia się podpisu, to nie jest on w stanie fałszywie zaprzeczyć swojemu autentycznemu, poprawnemu podpisowi.

Podpisy niezaprzeczalne mogą być stosowane w różnych sytuacjach (na podstawie [2; 4]). Jedną z możliwości zastosowania jest np. ochrona prywatności osoby – sygnowanie podpisem niezaprzeczalnym informacji przekazanej danemu podmiotowi (np. prasie) uniemożliwia sprawdzenie autentyczności wiadomości osobom nieuprawnionym. Inną możliwością wykorzystania podpisów niezaprzeczalnych jest składanie ich na pakietach oprogramowania – sprzedając taki pakiet, podmiot sprzedający przeprowadza transakcję z konkretnym nabywcą, który może sprawdzić autentyczność oprogramowania; gdyby jednak nabywca zrobił kopię i sprzedał pakiet kolejnemu nabywcy, ten trzeci nie może zweryfikować autentyczności bez udziału w tym procesie podmiotu pierwotnie sprzedającego (pierwszy kupujący traci wiarygodność i jest „na cenzurowanym"). Kolejną możliwość zastosowania podpisu niezaprzeczalnego opisuje następująca sytuacja: klient posiada w banku skrzynkę depozytową; każdorazowe skorzystanie ze skrzynki musi być przez klienta potwierdzone, np. datą i godziną. Jeśli klient złoży na potwierdzeniu podpis niezaprzeczalny, bank nie jest w stanie nikomu udowodnić, bez skontaktowania się z właścicielem skrzynki (klientem), że ten ze skrzynki skorzystał.

Konstruowanie schematów pozwalających na składanie podpisów niezaprzeczalnych jest możliwe dzięki technikom kryptograficznym.

Schemat podpisu niezaprzeczalnego jest podzielony na etapy, a głównymi podmiotami są podmiot podpisujący i weryfikujący. Postać (wartość) podpisu niezaprzeczalnego jest uzależniona od dokumentu, pod którym ma być złożony, oraz od klucza prywatnego podmiotu podpisującego. Można wyszczególnić następujące etapy protokołu podpisu niezaprzeczalnego (na podstawie [1]):

1. Generowanie klucza – podmiot podpisujący generuje dwa klucze: podpisywania i weryfikacji. Pierwszy z nich jest kluczem prywatnym (tajnym), natomiast drugi z nich jest kluczem publicznym.
2. Generowanie podpisu – mając do dyspozycji dokument oraz klucz podpisywania, podmiot podpisujący oblicza podpis pod tym dokumentem.
3. Weryfikacja podpisu – dokonuje się jej z użyciem dokumentu, podpisu i klucza weryfikacji podmiotu podpisującego. Strona weryfikująca wysyła podpisującemu zapytanie. Podpisujący, posiadając dokument, zapytanie i tajny klucz, oblicza odpowiedź i przesyła ją podmiotowi weryfikującemu. Dysponując dokumentem, podpisem, kluczem weryfikacji podpisującego oraz otrzymaną odpowiedzią na zapytanie, weryfikujący przeprowadza weryfikację i podpis zostaje zaakceptowany albo odrzucony.

3. Schematy Chauma-van Antwerpena

3.1. Schemat Chauma-van Antwerpena podpisu niezaprzeczalnego

Bezpieczeństwo przedstawianego algorytmu wynika z trudności rozwiązywania zagadnienia znajdowania logarytmu dyskretnego w cyklicznej podgrupie rzędu q należącej do $\mathbf{Z}_p^*$ (według [2]):

1. Generuje się klucz prywatny i publiczny podpisującego – wybiera się liczby pierwsze p i q takie, że q dzieli $p - 1$. Wybiera się generator g grupy cyklicznej rzędu q w $\mathbf{Z}_p^*$: losuje się element $b \in \mathbf{Z}_p^*$ i oblicza wartość $g = b^{(p-1)/q} \text{MOD } p$, przy czym jeżeli $g = 1$, należy wylosować inną wartość $b \in \mathbf{Z}_p^*$, taką by $g \neq 1$. Następnie wybiera się losowo liczbę pierwszą a, $a \in (1, q - 1]$ i oblicza wartość $y = g^a \text{MOD } p$. Kluczem prywatnym podmiotu podpisującego jest a, natomiast kluczem publicznym jest czwórka (p, q, g, y).
2. Podpis s, generowany pod wiadomością (dokumentem) w przyjmuje postać $s = w^a \text{MOD } p$.
3. Weryfikacja podpisu przebiega etapami: podmiot weryfikujący otrzymuje klucz publiczny podpisującego, losuje tajne niezerowe wartości całkowite $u, v \in \mathbf{Z}_q$ i oblicza wartość zapytania $z = s^u y^v \text{MOD } p$, którą wysyła do podpisującego. Podmiot podpisujący oblicza $o = (z)^{a^{-1}} \text{MOD } p$, gdzie a^{-1} jest odwrotnością do a modulo q – odpowiedź jest odsyłana podmiotowi weryfikującemu, ten z kolei, oblicza wartość $o' = w^u g^v \text{ MOD } p$ i jeśli wartość ta jest równa wartości o – podpis jest akceptowany, w innym przypadku – odrzucany.

Poprawność powyższego schematu wynika z następujących przekształceń:

$$o = (z)^{a^{-1}} \text{MOD } p = (s^u y^v)^{a^{-1}} \text{MOD } p = (w^{au} g^{av})^{a^{-1}} \text{MOD } p =$$
$$= (w^u g^v)^{aa^{-1}} \text{MOD } p = w^u g^v \text{ MOD } p = z' \text{MOD } p = o'$$

Jeżeli podpis s jest sfałszowany, prawdopodobieństwo tego, że podmiot weryfikujący nie odrzuci podpisu, wynosi co najwyżej $\frac{1}{q}$, co w przypadku dostatecznie dużej wartości q jest małą liczbą.

Przykład 1

Niech $p = 47$ i $q = 23$ spełniają warunek, że q jest dzielnikiem $p - 1$. Dla wybranej wartości $b = 13 \in \mathbf{Z}_{47}^*$ generatorem grupy cyklicznej rzędu 23 w $\mathbf{Z}_{47}^*$ jest:

$$g = b^{(p-1)/q} \text{MOD } p = 13^2 \text{ MOD } 47 = 28 .$$

Niech dalej całkowita liczba a, $a \in (1, 22]$ jest równa $a = 11$. Wówczas wartość y wynosi:

$$y = g^a \text{MOD } p = 28^{11} \text{ MOD } 47 = 18 .$$

Kluczem publicznym podpisującego jest czwórka (47, 23, 28, 18), a jego kluczem prywatnym wartość $a = 11$.

Niech wiadomością w, pod którą ma być złożony podpis, jest $w = 110101$. Podpisem s, składanym pod wiadomością w, będzie wartość:

$$s = w^a \text{MOD } p = 110101^{11} \text{MOD } 47 = 17 .$$

Niech wartości $u, v \in \mathbf{Z}_q$, które losuje podmiot weryfikujący po uzyskaniu klucza publicznego, wynoszą $u = 9$, $v = 15$. Wartość zapytania przesyłana następnie do składającego podpis wynosi:

$$z = s^u y^v \text{ MOD } p = 17^9 \cdot 18^{15} \text{ MOD } 47 = 25 .$$

Podpisujący oblicza odpowiedź, uzyskując wartość:

$$o = (z)^{a^{-1}} \text{ MOD } p = 25^{21} \text{ MOD } 47 = 37 ,$$

gdzie $a^{-1} = 21$ jest odwrotnością do $a = 11$ modulo 23. Wartość odpowiedzi jest wysyłana podmiotowi weryfikującemu, który dokonuje obliczenia:

$$o' = w^u g^v \text{ MOD } p = 110101^9 28^{15} \text{ MOD } 47 = 37 .$$

Ponieważ uzyskano równość wartości $o = o'$, więc podpis jest akceptowany.

Podmiot składający podpis, w jakiejś niedogodnej dla siebie sytuacji, mógłby chcieć wyprzeć się swojego poprawnego podpisu. Istnieje jednak możliwość sprawdzenia, za pomocą odpowiedniego algorytmu zabezpieczającego przed wyparciem, czy podmiot próbuje wyprzeć się ważnego podpisu; algorytm ten umożliwia także sprawdzenie, czy podpis nie jest sfałszowany.

Protokół Chauma-van Antwerpena chroniący przed wyparciem się podpisu

Protokół ten, dla danego podpisu *s* oraz klucza publicznego podmiotu podpisującego, wymaga m.in. dwukrotnego zastosowania schematu weryfikacji opisanego wyżej (na podstawie [2]):

1. Podmiot weryfikujący, po uzyskaniu klucza publicznego podmiotu podpisującego, losuje tajne niezerowe wartości całkowite $u_1, v_1 \in \mathbf{Z}_q$ i oblicza wartość $z = s^{u_1} y^{v_1} \text{ MOD } p$, którą przesyła podpisującemu. Podmiot podpisujący oblicza $o = (z)^{a^{-1}} \text{ MOD } p$ i odsyła ją podmiotowi weryfikującemu. Jeśli wartość *o* jest równa wartości wyznaczonej według wzoru $o = w^{u_1} g^{v_1} \text{ MOD } p$, podpis *s* jest akceptowany i protokół jest zatrzymywany.
2. Ponownie podmiot weryfikujący losuje tajne niezerowe wartości całkowite $u_2, v_2 \in \mathbf{Z}_q$ i oblicza wartość $z' = s^{u_2} y^{v_2} \text{ MOD } p$, którą przesyła podpisującemu. Podmiot podpisujący oblicza $o' = (z')^{a^{-1}} \text{ MOD } p$ i odsyła ją podmiotowi weryfikującemu. Jeśli zachodzi $o' = w^{u_2} g^{v_2} \text{ MOD } p$, podpis *s* jest akceptowany i protokół jest zatrzymywany.
3. Podmiot weryfikujący oblicza wartości $c = (og^{-v_1})^{u_2} \text{ MOD } p$ oraz $c' = (o' g^{-v_2})^{u_1} \text{ MOD } p$. Jeżeli wartości te są równe, wnioskuje się, że podpis *s* jest sfałszowany. W przeciwnym wypadku uznaje się, że podpis jest ważny, a podmiot podpisujący usiłuje się podpisu wyprzeć.

Przykład 2

Niech wartości niezbędne do złożenia i zweryfikowania podpisu są takie same, jak w przykładzie 1:

$p = 47$, $q = 23$,
$b = 13$, $g = 28$,
$a = 11$, $y = 18$.

Kluczem publicznym podpisującego jest czwórka (47, 23, 28, 18), a kluczem prywatnym wartość $a = 11$. Wiadomością *w* jest wartość $w = 110101$. Podpis, wyznaczony na podstawie schematu Chauma-van Antwerpena, wynosi 17. Założono jednak, że niepowołana osoba sfałszowała podpis i wysyła weryfikującemu wartość $s = 16$. W celu wykazania fałszerstwa jest przeprowadzany protokół Chauma-van Antwerpena chroniący przed wyparciem się podpisu.

Weryfikujący losuje niezerowe wartości całkowite $u_1, v_1 \in \mathbf{Z}_{23}$. Niech wartości te wynoszą $u_1 = 9, v_1 = 4$. Wartość zapytania wysyłana do składającego podpis wynosi:

$$z = s^{u_1} y^{v_1} \text{ MOD } p = 16^9 \cdot 18^4 \text{ MOD } 47 = 21.$$

Podpisujący oblicza następującą odpowiedź, którą odsyła weryfikującemu:

$$o = (z)^{a^{-1}} \text{ MOD } p = 21^{21} \text{ MOD } 47 = 34,$$

gdzie $a^{-1} = 21$ jest odwrotnością do a = 11 modulo 23. Podmiot weryfikujący porównuje tę wartość z odpowiadającą jej wartością obliczoną według wzoru:

$$o = w^{u_1} g^{v_1} \text{ MOD } p = 110101^9 28^4 \text{ MOD } 47 = 36.$$

Ponieważ wartości są różne, więc jest przeprowadzany kolejny etap protokołu.

Weryfikujący losuje kolejne niezerowe wartości całkowite $u_2, v_2 \in \mathbf{Z}_{23}$. Niech $u_2 = 8, v_2 = 5$. Wartość zapytania wysyłana do składającego podpis wynosi tym razem:

$$z' = s^{u_2} y^{v_2} \text{ MOD } p = 16^8 \cdot 18^5 \text{ MOD } 47 = 6.$$

Odpowiedź, którą ponownie oblicza podpisujący i odsyła weryfikującemu, wynosi:

$$o' = (z)^{a^{-1}} \text{ MOD } p = 6^{21} \text{ MOD } 47 = 17.$$

Podmiot weryfikujący porównuje tę wartość z wartością obliczoną za pomocą wzoru:

$$o' = w^{u_2} g^{v_2} \text{ MOD } p = 110101^8 28^5 \text{ MOD } 47 = 6.$$

Ponieważ wartości są różne, więc jest przeprowadzany ostatni etap protokołu. Podmiot weryfikujący oblicza elementy *c* i *c*':

$$c = (og^{-v_1})^{u_2} \text{ MOD } p = (34 \cdot 28^{-4})^8 \text{ MOD } 47 = (34 \cdot 14)^8 \text{ MOD } 47 = 24,$$

gdzie 14 jest odwrotnością do 28^4 modulo 47 oraz:

$$c' = (o' g^{-v_2})^{u_1} \text{ MOD } p = (17 \cdot 28^{-5})^9 \text{ MOD } 47 = (17 \cdot 24)^9 \text{ MOD } 47 = 24,$$

gdzie 24 jest odwrotnością do 28^5 modulo 47. Ponieważ $c = c'$, więc na podstawie protokołu wnioskuje się, że podpis s = 16 jest sfałszowany.

4. Przekształcalne podpisy niezaprzeczalne

Podpis przekształcalny można weryfikować, zaprzeczać oraz przekształcać do postaci zwykłego podpisu cyfrowego. Przy jego konstrukcji wykorzystuje się algorytm podpisu ElGamala. Algorytm przekształcalnego podpisu niezaprzeczalnego przebiega następująco (na podstawie [4]):

1. Wybiera się liczby pierwsze p i q takie, że q dzieli $p - 1$, wybiera się generator g grupy cyklicznej rzędu q w $\mathbf{Z}_p^*$: losuje się element $b \in \mathbf{Z}_p^*$ i oblicza wartość $g = b^{(p-1)/q} \text{MOD } p$, przy czym jeżeli $g = 1$, należy wylosować inną wartość $b \in \mathbf{Z}_p^*$ taką, by $g \neq 1$. Wybiera się losowo liczby a i d, takie że $a, d \in (1, q - 1]$ i oblicza wartości $y = g^a \text{MOD } p$ oraz $x = g^d \text{MOD } p$. Kluczami prywatnymi podmiotu podpisującego są a i d, natomiast kluczem publicznym jest piątka liczb (p, q, g, y, x).
2. W celu obliczenia przekształcalnego podpisu cyfrowego składanego pod wiadomością w losuje się niezerową liczbę $t \in \mathbf{Z}_q$ i oblicza się wartości $T = g^t \text{MOD } p$ oraz $w' = Ttdw \text{MOD } q$.
3. Generuje się zwykły podpis ElGamala dla wiadomości w' – losuje się liczbę k, $k \leq p - 1$ taką, że NWD$(k, q - 1) = 1$. Oblicza się wartość l według reguły $l = g^k \text{MOD } p$ i za pomocą algorytmu Euklidesa wyznacza się podpis s z zależności:

$$w' \equiv la + ks \pmod{q}.$$

   Wartości (l, s) oraz T stanowią podpis ElGamala.
4. Podmiot podpisujący sprawdza swój podpis i pokazuje go weryfikującemu: podmiot weryfikujący losuje dwie liczby $\delta, \varepsilon \in \mathbf{Z}_p$, oblicza $c = T^{Tw\delta} g^{\varepsilon} \text{ MOD } p$ i przesyła wartość podpisującemu. Podpisujący wybiera liczbę losową γ i oblicza $h_1 = cg^{\gamma} \text{MOD } p$ oraz $h_2 = h_1^d \text{MOD } p$. Wartości te są wysyłane weryfikującemu, a ten przesyła podpisującemu wartości δ i ε. Podpisujący sprawdza równość $c = T^{Tw\delta} g^{\varepsilon} \text{ MOD } p$ i przesyła γ weryfikującemu. Ten z kolei sprawdza, czy zachodzi $h_1 = T^{Tw\delta} g^{\varepsilon+\gamma} \text{MOD } p$ oraz $h_2 = y^{l\delta} l^{s\delta} x^{\varepsilon+\gamma} \text{MOD } p$.

   Poprawność powyższego schematu wynika z następujących przekształceń:

$$h_1 = T^{Tw\delta} g^{\varepsilon+\gamma} \text{MOD } p = cg^{\gamma} \text{MOD } p = h_1$$

$$h_2 = y^{l\delta} l^{s\delta} x^{\varepsilon+\gamma} \text{MOD } p = g^{al\delta} g^{ks\delta} g^{d(\varepsilon+\gamma)} \text{MOD } p = \left(g^{la+ks}\right)^{\delta} \left(g^{\varepsilon+\gamma}\right)^{d} \text{MOD } p =$$

$$= \left(g^{w'}\right)^{\delta} \left(g^{\varepsilon+\gamma}\right)^{d} \text{MOD } p = \left(g^{Ttdw}\right)^{\delta} \left(g^{\varepsilon+\gamma}\right)^{d} \text{MOD } p =$$

$$= \left(g^{tTw\delta}\right)^{d} \left(g^{\varepsilon+\gamma}\right)^{d} \text{MOD } p = \left(T^{Tw\delta} g^{\varepsilon+\gamma}\right)^{d} \text{MOD } p = h_1^{d} \text{ MOD } p = h_2 .$$

Jeśli zostanie powszechnie udostępniona wartość *d*, podpisujący może przekształcić swoje podpisy niezaprzeczalne w zwykłe podpisy cyfrowe, a zatem takie, których weryfikacja nie wymaga interakcji z podpisującym.

Przykład 3

Niech początkowe wartości umożliwiające złożenie podpisu są takie same, jak w przykładzie 1, zatem:
$p = 47$, $q = 23$,
$b = 13$, $g = 28$,
$a = 11$, $y = 18$.

Ponadto niech $d = 17$, wówczas:

$$x = g^{d} \text{MOD } p = 28^{17} \text{ MOD } 47 = 21 .$$

Kluczem publicznym podpisującego jest piątka (47, 23, 28, 18, 21), a kluczem prywatnym są wartości $a = 11$ oraz $d = 17$. Wiadomością, którą podpisujący chce zatwierdzić, jest wartość $w = 110$.

Obliczenie podpisu wymaga wylosowania różnej od zera liczby $t \in \mathbf{Z}_{23}$; niech $t = 19$. Oblicza się wartości:

$$T = g^{t} \text{MOD } p = 28^{19} \text{ MOD } 47 = 14 ,$$

$$w' = Ttdw \text{ MOD } q = 14 \cdot 19 \cdot 17 \cdot 110 \text{ MOD } 23 = 22 .$$

Generowany jest podpis ElGamala dla wiadomości *w*'. Niech liczba k, $k \leq 46$ oraz względnie pierwsza z $q - 1 = 22$ jest równa $k = 41$. Wówczas wartość l wynosi:

$$l = g^{k} \text{MOD } p = 28^{41} \text{ MOD } 47 = 24 ,$$

a podpis *s* wyznacza się z zależności $w' \equiv la + ks \pmod{q}$, zatem:

$$22 \equiv 24 \cdot 11 + 41 \cdot s \pmod{23} ,$$

skąd $s = 7$. Podpisem jest para (24, 7) oraz wartość $T = 14$.

Weryfikacja podpisu dokonuje się etapami. Podmiot weryfikujący losuje dwie liczby: δ i ε; niech $\delta = 7$ oraz $\varepsilon = 12$. Wykorzystując te liczby, oblicza wartość:

$$c = T^{Tw\delta} g^{\varepsilon} \text{ MOD } p = 14^{14 \cdot 110 \cdot 7} 28^{12} \text{ MOD } 47 = 21,$$

którą przesyła podpisującemu. Ten z kolei losuje liczbę γ; niech γ = 16. Podpisujący oblicza dwie wartości:

$$h_1 = cg^{\gamma} \text{MOD } p = 21 \cdot 28^{16} \text{ MOD } 47 = 4,$$

$$h_2 = h_1^d \text{MOD } p = 4^{17} \text{ MOD } 47 = 27.$$

Wartości te wysyła weryfikującemu, a ten przesyła podpisującemu wartości δ i ε. Za pomocą tych wartości podpisujący sprawdza równość:

$$c = T^{Tw\delta} g^{\varepsilon} \text{ MOD } p$$

i przesyła γ weryfikującemu. Ostatecznie weryfikujący sprawdza, czy zachodzi:

$$h_1 = T^{Tw\delta} g^{\varepsilon+\gamma} \text{MOD } p = 14^{14 \cdot 110 \cdot 7} 28^{12+16} \text{MOD } 47 = 4$$

oraz:

$$h_2 = y^{l\delta} l^{s\delta} x^{\varepsilon+\gamma} \text{MOD } p = 18^{24 \cdot 7} 24^{7 \cdot 7} 21^{12+16} \text{MOD } 47 = 27.$$

Podsumowanie

Bezpieczeństwo implementowanych protokołów, a także ich funkcjonalność są jednymi z najistotniejszych zagadnień, jakimi zajmuje się współczesna kryptografia. Tworzenie bezpiecznych algorytmów jest możliwe dzięki stosowaniu pewnych trudnych do rozwiązania problemów dla dużych liczb całkowitych, takich jak np. znajdowanie logarytmu dyskretnego w cyklicznej podgrupie rzędu q należącej do $\mathbf{Z}_p^*$. Trudność rozwiązywania tych zagadnień daje gwarancję bezpieczeństwa wykorzystywanych protokołów, a zarazem bezpieczeństwa przeprowadzanych operacji.

Przedstawione w tej pracy protokoły Chauma-van Antwerpena oraz protokół przekształcalnych podpisów niezaprzeczalnych są protokołami podpisów ukazującymi zastosowanie zmodyfikowanych algorytmów podpisów cyfrowych, których weryfikacja wymaga bezpośredniej konfrontacji z podmiotem składającym podpis. Jest to użyteczne zwłaszcza wtedy, gdy podmiot podpisujący chce mieć kontrolę nad weryfikacją podpisu oraz nad podmiotami weryfikującymi.

Istnieją inne warianty podpisów niezaprzeczalnych (według [2; 4]). W niektórych rozgranicza się relacje: podpisujący – wiadomość i podpisujący – podpis – istnieje wówczas możliwość sprawdzenia przez dowolny podmiot, czy podpisujący złożył dany podpis, a konfrontacja z podpisującym jest wymagana

tylko w celu zweryfikowania, czy dany podpis jest właściwy dla konkretnej wiadomości. Pewnym wariantem podpisów niezaprzeczalnych są także powierzone podpisy niezaprzeczalne, które charakteryzują się tym, iż protokół zaprzeczający (wykonywany w przypadku wątpliwości dotyczących podpisu) może być przeprowadzony przez trzecią, niezależną stronę. Nieco innym wariantem są podpisy cyfrowe z wyznaczonym potwierdzającym – w tym wariancie weryfikacja podpisu może być dokonana przez osobę podpisującą, a także przez trzeci, specjalnie do tego celu wyznaczony podmiot. Takie podpisy mogą zapobiegać fałszywym zastosowaniom podpisu, stanowią ochronę w przypadku utraty klucza, dają możliwość zweryfikowania podpisu, gdy podmiot podpisujący jest aktualnie nieobecny. Jeszcze inną wersję podpisów można stworzyć łącząc niezaprzeczalne podpisy cyfrowe z algorytmami podziału sekretu – istniałaby wówczas możliwość scedowania weryfikacji podpisu na wybraną grupę n osób, z której do każdego protokołu weryfikacji byłaby potrzebna podgrupa złożona z co najmniej k, $k \leq n$, osób.

Można zauważyć, że możliwości zastosowań omawianych zagadnień są dosyć duże, a ich użyteczność znacząca. Tworzenie i stosowanie protokołów nie byłoby jednak możliwe bez implementacji pojęć matematycznych. Zatem znajomość pojęć, zwłaszcza z zakresu teorii liczb i algebry, jest kluczowym elementem współtworzenia współczesnej kryptografii.

Literatura

Wydawnictwa zwarte

1. Buchmann A.J.: *Wprowadzenie do kryptografii*. Wydawnictwo Naukowe PWN, Warszawa 2006.
2. Menezes A.J., Oorschot P.C. van, Vanstone S.A.: *Kryptografia stosowana*. Wydawnictwo Naukowo-Techniczne, Warszawa 2005.
3. Ross K.A., Wright C.R.B.: *Matematyka dyskretna*. Wydawnictwo Naukowe PWN, Warszawa 2003.
4. Schneier B.: *Kryptografia dla praktyków*. Wydawnictwo Naukowo-Techniczne, Warszawa 2002.
5. Song Y. Yan: *Teoria liczb w informatyce*. Wydawnictwo Naukowe PWN, Warszawa 2006.

Strony internetowe

http://www.podpis.nordea.pl.

UNDENIABLE SIGNATURES – SIGNATURES WITH ADDITIONAL FUNCTIONALITY

Summary

Today, the role of digital signatures becoming more and more significant. The signature provides authentication and integrity of a message but quite often other features like additional functionality is required from the signature. Because of that the algorithms used in ordinary signing are not sufficient, they need to be special algorithms with some additional characteristics. Undeniable signatures are a form of digital signature that have two distinctive features: the verification process is interactive (the signatures cannot be verified without signer's cooperation), a disavowal protocol allows to determine whether a given signature is a forgery. These signatures can be used, for example, to protect a software against unauthorized persons. The safety of these signatures is guaranteed by techniques of modern cryptography which is based on advanced mathematical tools and computer technology.

The purpose of the article is to present selected schemes (protocols) of undeniable signatures (based on discrete log systems) such as Chaum-van Antwerpen protocol, Chaum-van Antwerpen protocol protecting against disavowal and a convertible undeniable signature protocol. Main definitions and theorems are presented and all described protocols are illustrated with examples.

Zygmunt Przybycin

DWUMIANOWY MODEL WYCENY OPCJI W WARUNKACH ROZMYTYCH INFORMACJI

Wprowadzenie

Inwestowanie na rynku kapitałowym jest sztuką i wyzwaniem dla osób pragnących osiągać ponadprzeciętne zyski. Sztuka polega na takim zarządzaniu inwestycjami kapitałowymi, które zapewnia osiągnięcie założonego celu. Tym celem może być minimalizacja ryzyka inwestycyjnego. Zarówno w literaturze, jak i w praktyce spotyka się wiele narzędzi i metod wspomagających proces zarządzania ryzykiem inwestycyjnym. Jedną z metod zarządzania ryzykiem inwestycyjnym jest przenoszenie ryzyka na inne podmioty. Narzędziem umożliwiającym przenoszenie ryzyka są kontrakty terminowe, a w szczególności opcje. Opcje w początkowym okresie były stosowane jako instrument finansowy, który zabezpieczał inwestycję przed ryzykiem, lecz bardzo szybko zorientowano się, że możliwości tego instrumentu są o wiele wieksze. Opcje zaczęto również stosować jako instrument finansowy generujący ponadprzeciętne zyski przy relatywnie niskim zaangażowaniu kapitału. To szerokie zastosowanie opcji spowodowało, iż stały się przedmiotem zainteresowania inwestorów. W historii inwestowania są znane spektakularne sukcesy stosowania tego instrumentu, a także liczne porażki inwestycji z wykorzystaniem opcji. Wydaje się, że dużego znaczenia w inwestycjach z zastosowaniem opcji nabiera problem oszacowania wartości opcji. Chodzi mianowicie o określenie tzw. wartości sprawiedliwej. W literaturze są znane rożne modele wyceny opcji. W tym artykule podjęto próbę modyfikacji dwumianowego modelu wyceny opcji. Celem jest zaadoptowanie dwumianowego modelu wyceny opcji w warunkach rozmytych informacji rynkowych.

1. Opcje – wybrane pojęcia

Ważną grupą instrumentów finansowych rynku kapitałowego są instrumenty pochodne, które zostały „wymyślone" w celu zarządzania wartością oraz ryzykiem inwestycji kapitałowych. Instrumenty te z uwagi na funkcje, jakie speł-

niają, podlegają ciągłemu rozwojowi i licznym modyfikacjom. Instrumentem pochodnym, który cieszy się dużym zainteresowaniem inwestorów, są opcje finansowe.

Opcja jest instrumentem finansowym, który daje jego posiadaczowi prawo do wykonania opcji, natomiast na wystawcę opcji nakłada obowiązek jej zrealizowania, tj. dostarczenia instrumentu bazowego posiadaczowi opcji w sytuacji, gdy zechce on ją zrealizować. Ze względu na sposób realizacji opcji wyróżnia się opcje kupna oraz sprzedaży. Opcja kupna daje jej posiadaczowi prawo do zakupu instrumentu bazowego po określonej cenie w określonym czasie, natomiast opcja sprzedaży daje takie same prawo, ale w odniesieniu do sprzedaży instrumentu bazowego. W kontrakcie opcyjnym występuje również wystawca opcji, który przyjmuje na siebie obowiązek dostarczenia instrumentu bazowego w przypadku, gdy opcja zostanie wykonana. Podstawowymi charakterystykami opcji są:

- cena wykonania (X) – cena, po jakiej opcja jest wykonana; jest ustalona w momencie wystawienia opcji i nie zmienia się w czasie ważności opcji,
- cena instrumentu bazowego (S_t) – wartość rynkowa instrumentu bazowego, na który jest wstawiona opcja,
- cena opcji (inaczej premia) (C) – cena prawa, które nabywa posiadacz opcji; cena ta jest kształtowana przez rynek,
- termin wygaśnięcia opcji (T) – termin ważności opcji, po upływie którego opcja traci ważność i nie może być wykonana,
- termin wykonania opcji (t_o) – termin, w którym opcja jest wykonana.

W zależności od możliwości wykonania opcji wyróżnia się opcję amerykańską oraz opcję europejską. Opcję amerykańską można wykonać w dowolnym momencie od chwili nabycia do momentu wygaśnięcia ($t_0 \leq$ T), natomiast opcję europejską można wykonać tylko w momencie wygaśnięcia (t_0= T). W dalszej części artykułu będą omawiane tylko opcje europejskie. Posiadacz opcji, jak wspomniano, ma prawo do jej wykonania i wykona opcję tylko wówczas, gdy będzie to opłacalne – decyduje relacja pomiędzy ceną wykonania a ceną instrumentu bazowego. Wyróżnia się tutaj następujące sytuacje:

- opcję opłaca się wykonać – opcja jest w cenie,
- opcji nie opłaca się wykonać – opcja nie jest w cenie,
- opcja jest neutralna – opcja jest po cenie.

Dla opcji kupna: opcja jest w cenie, gdy cena wykonania jest niższa od ceny instrumentu bazowego, w przeciwnym wypadku opcja nie jest w cenie, jeżeli natomiast cena wykonania opcji jest równa cenie instrumentu bazowego, to mówi się, że opcja jest neutralna. Dla opcji sprzedaży zachodzi odwrotna relacja pomiędzy ceną wykonania a ceną instrumentu bazowego.

Innym ważnym parametrem charakteryzującym opcje bez względu na rodzaj instrumentu bazowego jest wartość opcji, czyli wypłata, jaką otrzyma posiadacz opcji w chwili jej wykonania. Wartość opcji jest sumą dwóch składowych: wartości wewnętrznej oraz wartości czasowej zwanej również wartością zewnętrzną. Wartość opcji można więc zapisać następująco:

$$W = W_w + W_t \tag{1}$$

gdzie:
W_w – wartość wewnętrzna opcji,
W_t – wartość czasowa opcji.

Opcja ma dodatnią wartość wewnętrzną, gdy jest w cenie, w przeciwnym wypadku wartość ta jest równa zero. W szczególności dla opcji kupna wartość wewnętrzna opcji wyraża się wzorem:

$$W_w = \max \{S_t - X, 0\} \tag{2}$$

natomiast dla opcji sprzedaży wzorem:

$$W_w = \max \{X - S_t, 0\} \tag{3}$$

Z zależności (1) wynika, że wartość czasowa opcji maleje do zera wraz ze zbliżaniem się terminu wykonania do terminu wygaśnięcia. Wypada również zauważyć, iż na wartość opcji mają wpływ następujące czynniki:
– cena wykonania opcji,
– cena instrumentu bazowego,
– długość okresu ważności opcji,
– zmienność stopy zwrotu instrumentu bazowego,
– stopa wolna od ryzyka.

W przypadku opcji kupna cena wykonania jest destymulantą, natomiast pozostałe czynniki są stymulantami wartości opcji. W przypadku opcji sprzedaży destymulantami są: cena instrumentu pierwotnego oraz stopa wolna od ryzyka, zaś pozostałe czynniki są stymulantami.

Wartość opcji, zwana również ceną sprawiedliwą jako wypadkowa czynników rynkowych, zmienia się w czasie ważności opcji i pokrywa się z ceną opcji tylko w przypadku, gdy rynek jest efektywny. Dlatego też ważne z punktu widzenia praktyki jest w miarę precyzyjne oszacowanie wartości opcji. Wartość sprawiedliwą opcji szacuje się wykorzystując różne modele wyceny. W artykule tym ograniczono się do modelu dwumianowego wyceny wartości opcji przy założeniu nieostrych informacji kształtowania się cen instrumentu bazowego.

2. Logika rozmyta i jej zastosowanie w modelu wyceny opcji

Proces szacowania prawdziwej wartości opcji z uwagi na stosunkowo dużą ilość informacji jest procesem złożonym. Ponadto część informacji ma charakter prognostyczny, a więc nie są to informacje w pełni precyzyjne. Oznacza to, że w procesie szacowania wartości opcji występuje tzw. zasada niespójności, według której w modelowaniu złożonych systemów stosuje się obok informacji precyzyjnych również informacje nieprecyzyjne (nieostre). Dalej informacje nieprecyzyjne występujące w modelu wyceny opcji będą traktowane w kategoriach liczb rozmytych.

Liczbą rozmytą nazywa się zbiór rozmyty określony na przestrzeni liczb rzeczywistych. Formalnie liczbę rozmytą A określa się następująco:

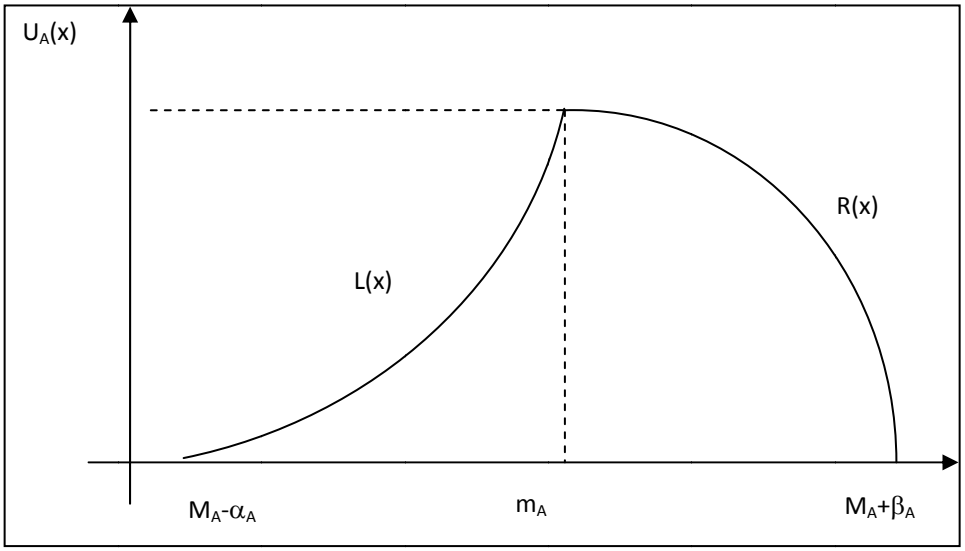
$$A = \{(U_A(x), x)\} \ \forall x \in R \tag{4}$$

gdzie $U_A(x)$ jest funkcją przynależności zbioru rozmytego, która każdej liczbie $x \in R$ przypisuje stopień jej przynależności do zbioru A, przy czym $U_A(x) \in <0,1>$.

Liczbę rozmytą jednoznacznie określa funkcja przynależności, którą wyznacza się na podstawie wiedzy historycznej lub/i eksperckiej. Ograniczono się tu tylko do trójkątnych liczb rozmytych L-R. Funkcję przynależności rozmytej trójkątnej liczby L-R określa się wzorem:

$$\begin{cases} L(x) \text{ dla } x \in (M_A - \alpha_{A,,} M_a) \\ U_A(x) = R(x) \text{ dla } x \in (M_A, M_A + \beta_A) \\ 1 \text{ dla } x = M_A \\ 0 \text{ dla x pozostałych} \end{cases} \tag{5}$$

gdzie:
L(x) – funkcja niemalejąca,
R(x) – funkcja nierosnąca.



Rys. 1. Rozmyta liczba trójkątna L-R

Przyjęto, że rozmyta liczba trójkątna L-R będzie zapisywana w postaci:

$$A = (M_A, \alpha_A, \beta_A) \tag{6}$$

Postać (6) nazywa się reprezentacją L-R liczby rozmytej. Przykładowo A = (2; 0,5; 0,8) oznacza rozmytą liczbę trójkątną około 2. W szczególności jeżeli $\alpha_A = \beta_A = 0$, wówczas liczbę rozmytą (6) traktuje się jak liczbę ostrą. Używając zapisu (6), definiuje się liczbę przeciwną do liczby rozmytej A:

$$-A = (-M_A, \alpha_A, \beta_A)$$

oraz liczbę odwrotną:

$$A^{-1} = (1/M_A, \beta_A / M_A (M_A + \beta_A), \alpha_A / M_A (M_A - \alpha_A))$$

przy założeniu $M_A \neq 0$.

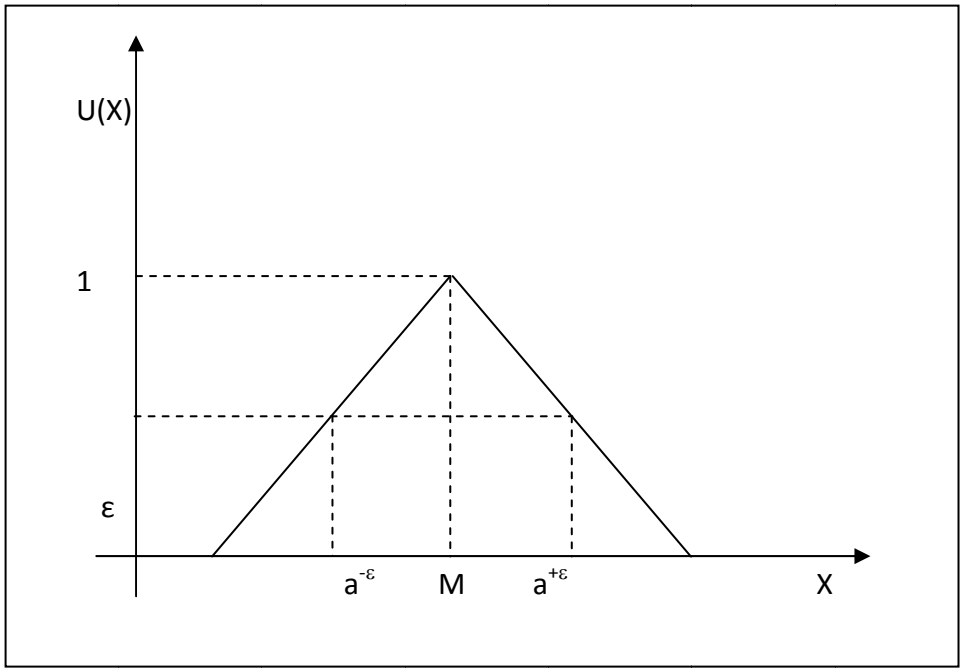
Wprowadzono pojęcie przekroju liczby rozmytej: ε przekrojem liczby rozmytej (4) nazywa się zbiór ostry określony następująco:

$$A^{\varepsilon} = \{ x \in R : U_A(x) \geq \varepsilon \}, \varepsilon \in (0, 1] \tag{7}$$

Zbiór ten będzie również oznaczany symbolem $< a^{-\varepsilon}, a^{+\varepsilon} >$, gdzie:

$$a^{-\varepsilon} = \min \{ x : U_A(x) \geq \varepsilon \}, \tag{8}$$

$$a^{+\varepsilon} = \max \{ x : U_A(x) \geq \varepsilon \}. \tag{9}$$



Rys. 2. Liczba rozmyta i jej ε przekrój

Używając operacji ε przekroju, można dokonać dekompozycji funkcji przynależności liczby rozmytej zgodnie ze wzorem (zasada dekompozycji):

$$U_A(x) = \sup [\, \varepsilon \wedge I_{A^\varepsilon}(x) \,], \; x \in R, \; \varepsilon \in \langle 0, 1 \rangle \tag{10}$$

gdzie $I_{A^\varepsilon}(x)$ jest funkcją charakterystyczną zbioru A^ε, natomiast ^ – algebraiczną operacją minimum. Zasada dekompozycji wyraża więc funkcję przynależności liczby rozmytej poprzez funkcje charakterystyczne zbiorów ostrych. Na liczbach rozmytych definiuje się działania arytmetyczne. Ograniczono się tu do zdefiniowania dodawania i mnożenia rozmytych liczb trójkątnych typu L-R. [4].

Jeżeli liczby rozmyte A i B są liczbami typu L-R:

$$A = (M_A, \alpha_A, \beta_A), \; B = (M_B, \alpha_B, \beta_B) \tag{11}$$

to sumę liczb rozmytych A, B określa się następująco:

$$A+B = (M_{A+B}, \alpha_{A+B}, \beta_{A+B}) \tag{12}$$

gdzie:

$$M_{A+B} = M_A + M_B,$$
$$\alpha_{A+B} = (\alpha_A + \alpha_B),$$
$$\beta_{A+B} = (\beta_A + \beta_B).$$

Iloczyn dwóch liczb rozmytych A, B dla A > 0 oraz B > 0 określa wzór:

$$A \bullet B = (M_{A \bullet B}, \alpha_{A \bullet B}, \beta_{A \bullet B}) \tag{13}$$

gdzie:
$M_{A \bullet B} = M_A \bullet M_B$,
$\alpha_{A \bullet B} = M_A\ \alpha_B + M_B\ \alpha_A - \alpha_A \alpha_B$,
$\beta_{A \bullet B} = M_A\ \beta_B + M_B\ \beta_A + \beta_A\ \beta_B$.

Z przyjętych definicji sumy i iloczynu liczb rozmytych wynika, że własności dla tych działań są identyczne z własnościami analogicznych działań na liczbach ostrych. Funkcję przynależności dla tak zdefiniowanej sumy oraz iloczynu liczb rozmytych wyznacza się stosując zasadę dekompozycji [3]. W przypadku gdy funkcje L(.) oraz R(.) są funkcjami liniowymi, funkcje przynależności wyznacza się analitycznie.

Należy zwrócić uwagę na fakt, że prognozowane ceny instrumentu bazowego w modelu wyceny opcji różnią się od ceny rzeczywistej nawet wtedy, gdy są sporządzone według założeń, które są bliskie faktycznym wielkościom osiągniętym w przeszłości. Dlatego też prognozowana cena instrumentu bazowego powinna uwzględniać przynajmniej trzy warianty: pesymistyczny, najbardziej realny oraz optymistyczny. Warianty pesymistyczny oraz optymistyczny powinny obejmować takie wartości, które nie powinny być przekroczone w przyszłości, stąd będą one reprezentować planowane wartości nieprzekraczalne. Jeżeli prognozowana cena instrumentu bazowego jest określona w kategoriach liczb rozmytych, a w szczególności w postaci trójkątnej liczby rozmytej (6), to parametry występujące w zapisie rozmytej liczby A traktuje się odpowiednio:
M_A – wariant najbardziej realny,
$M_A - \alpha_A$ – wariant pesymistyczny,
$M_A + \beta_A$ – wariant optymistyczny.

Przyjęta interpretacja parametrów liczby rozmytej umożliwia również określenie rozmytego ryzyka wyceny opcji. Jeżeli założy się, że cena sprawiedliwa opcji jest wyrażona w postaci liczby rozmytej (6), to ryzyko rozmyte wyceny wartości opcji mierzy się następująco [5]:

$$S_A = (\alpha_A + \beta_A) / 2 \tag{14}$$

Przedział $< M_A - \alpha_A,\ M_A + \beta_A >$ jest ε przekrojem liczby rozmytej (6) (ε = 0). Przyjmując różne poziomy przekrojów ($\varepsilon \in \langle 0,1 \rangle$), można rozważać różne ε przekroje tej samej liczby rozmytej, a w szczególności różne scenariusze wyceny opcji.

3. Dwumianowy model wyceny opcji w warunkach rozmytych informacji

Zagadnienie wyceny instrumentów pochodnych jest zagadnieniem ważnym z punktu widzenia teorii i praktyki i trzeba dodać, że nadal aktualnym.

Pierwsze prace z zakresu metodologii wyceny instrumentów pochodnych pokazały się w 1973 roku i były to prace: F. Blacka i M. Scholesa (*The Pricing of Options annd Corporate Liabilities*) oraz R.C. Mertona (*Theory of Rational Option Pricing*). W pracach tych przedstawiono koncepcję arbitrażowej wyceny instrumentów pochodnych. Arbitrażowa koncepcja wyceny jest na ogół realizowana w konwencji portfela bez ryzyka lub w konwencji replikacji. Model wyceny opcji realizowany w konwencji portfela bez ryzyka został zaproponowany w 1972 roku przez J. Coxa, M. Rubinsteina i S. Rossa [1]. Koncepcja wyceny opcji w konwencji portfela bez ryzyka polega na takiej konstrukcji portfela posiadającego określoną ilość instrumentów bazowych oraz wystawionej opcji na ten instrument, aby był to portfel bez ryzyka.

Koncepcję portfela bez ryzyka prześledzono na przykładzie europejskiej opcji kupna na akcję bez dywidendy. Rozważono portfel, który składa się z *h* akcji oraz wystawionej na te akcje opcji kupna, przy czym wymaga się, aby inwestycja ta była bez ryzyka. Wartość portfela w chwili początkowej jest równa:

$$\mathrm{h}\, S_0 - C_0$$

gdzie :
S_0 – cena akcji w momencie początkowym inwestycji,
C_0 – wartość europejskiej opcji kupna w momencie początkowym inwestycji.

Założono dalej, że do terminu wygaśnięcia opcji pozostał jeden okres. Na koniec tego okresu cena akcji może wzrosnąć do poziomu S_u lub spaść do poziomu S_d. Zatem w terminie wykonania opcji wartość inwestycji jest odpowiednio równa:

$$V_u = \mathrm{h}\, S_u - C_u, \text{ gdy cena akcji wzrośnie}$$

$$V_d = \mathrm{h}\, S_d - C_d, \text{ gdy cena akcji spadnie}$$

przy czym wartość opcji, gdy cena akcji wzrośnie – $C_u = \max\{S_u - \mathrm{X}, 0\}$, natomiast wartość opcji, gdy cena akcji spadnie – $C_d = \max\{ S_d - \mathrm{X}, 0\}$, gdzie X oznacza cenę wykonania opcji.

Ponieważ założono, że inwestycja jest bez ryzyka, więc musi zachodzić równość:

$$V_u = V_d$$

Z równości tej wyznacza się współczynnik hedgingowy h:

$$h = (C_u - C_d) / (S_u - S_d)$$

Współczynnik hedgingowy określa liczbę akcji w portfelu, która zabezpiecza wystawioną opcję kupna. Jeżeli ponadto założy się, że mechanizm kształtowania ceny akcji w górę i w dół jest następujący:

$$S_u = u\, S_0 \text{ i } S_d = d\, S_0$$

przy czym $u > 1 + r_f > d$ (r_f – stopa zwrotu wolna od ryzyka), co wyklucza możliwość transakcji arbitrażowej, wówczas współczynnik hedgingowy wyraża się wzorem:

$$h = (C_u - C) / S_0(u - d) \tag{15}$$

Inwestycja bez ryzyka powinna generować stopę zwrotu równą stopie zwrotu wolnej od ryzyka, zatem zachodzi następująca równość:

$$(h\, S_u - C_u) / (h\, S_0 - C_0) = 1 + r_f$$

Wstawiając do powyższej równości współczynnik hedgingowy (15), otrzymano wartość euorepejskiej opcji kupna na akcję bez dywidendy:

$$C_0 = [g\, C_u + (1 - g)\, C_u] / (1 + r_f) \tag{16}$$

gdzie:

$$g = (1 + r_f - d) / (u - d) \tag{17}$$

Wobec wcześniejszego założenia $g \in (0, 1)$, tak więc para $(g, 1 - g)$ jest tzw. miarą arbitrażową [6].

Wzór (16) określa wartość europejskiej opcji kupna w przypadku jednego okresu, jaki pozostał do terminu wygaśnięcia opcji. Wzór ten można uogólnić na przypadek n okresów, jakie pozostały do terminu wygaśnięcia opcji ($n > 1$). Zakładając, że wielkości ruchu ceny akcji w górę (u) oraz w dół (d) są znane i nie zmieniają się w okresie ważności opcji, wówczas wartość sprawiedliwa opcji wyraża się wzorem [2]:

$$W_0 = \sum_{k=0}^{n} \binom{n}{k} g^k (1 - g)^{n-k} \max \{S_0\, u^k\, d^{n-k} - X, 0\} / (1 + r_f)^n \tag{18}$$

gdzie g jest określone wzorem (15), natomiast $\binom{n}{k}$ oznacza symbol Newtona.

Równość (18) określająca wartość europejskiej opcji kupna w literaturze jest nazywana dwumianowym modelem wyceny opcji. Zupełnie analogicznie wyprowadza się model europejskiej opcji sprzedaży. W dwumianowym modelu

wyceny opcji kupna zmienność ceny akcji jest zdeterminowana parametrami u oraz d. Parametry te określają zmienność ceny instrumentu bazowego w przyszłości, a więc mają charakter prognostyczny. Oznacza to, że nie są w pełni precyzyjne, stąd w modelu wyceny opcji ingeruje wspomniana wcześniej zasada niespójności. Zgodnie z tą zasadą w modelu wyceny opcji obok informacji precyzyjnych występują również informacje nieprecyzyjne – rozmyte. Założono więc, że parametry określające ruchy ceny instrumentu bazowego są liczbami rozmytymi, dla ułatwienia przyjęto, że są trójkątnymi liczbami rozmytymi typu L-R, ponadto założono znajomość funkcji przynależności tych liczb. Konsekwencją przyjętego założenia jest to, że wartość opcji w momencie wygaśnięcia jest również liczbą rozmytą. Wstawiając do modelu dwumianowego wyceny opcji kupna w miejsce parametrów określających ruchy ceny instrumentu bazowego ich wersje rozmyte, otrzymano model wyceny opcji w warunkach rozmytych informacji. Uwzględniając fakt, iż działania arytmetyczne na liczbach rozmytych mają podobne własności, jak działania na liczbach ostrych, otrzymano następujący model wyceny opcji kupna:

$$\breve{W}_0 = \sum_{k=0}^{n} \binom{n}{k} \breve{g}^{\,k} (1- \breve{g}\,)^{n-k} \max\{S\, \breve{u}^{\,k}\, \breve{d}^{\,n-k} - X\,,\, 0\}/(\,1 + r_f\,)^n \qquad (19)$$

gdzie symbol $\breve{W}$ oznacza liczbę rozmytą. Funkcję przynależności rozmytej wartości opcji kupna wyznacza się zgodnie z zasadą dekompozycji lub jej przybliżoną postać korzystając z reprezentacji L-R.

Podsumowanie

Zaproponowana w artykule modyfikacja modelu wyceny opcji powinna poprawić skuteczność wyznaczenia ceny sprawiedliwej. Stwierdzenie to można uzasadnić tym, że w naukach ekonomicznych bardziej realne i często wystarczające jest stwierdzenie, że np. cena instrumentu finansowego przyjmie wartość równą pewnej liczbie rozmytej, niż stwierdzenie, że wartość ta będzie określona liczbą ostrą. Należy wyraźnie zaznaczyć, iż parametry określające liczbę rozmytą wyznacza się na podstawie danych historycznych oraz wiedzy eksperckiej, dotyczy to zwłaszcza funkcji przynależności. Fakt ten pozwala mieć nadzieję na to, że w prognozach określających przyszłe wartości parametrów występujących w modelu wyceny opcji zostanie uwzględniona psychologia rynku. W zaprezentowanej modyfikacji modelu dwumianowego założono rozmytość parametrów określających ruchy ceny instrumentu bazowego, nic nie stoi więc na przeszkodzie, aby stopę wolną od ryzyka również traktować jako liczbę rozmytą. Ponadto przy stosowaniu logiki rozmytej istnieje możliwość oceny ryzyka oszacowanej sprawiedliwej wartości opcji.

Literatura

1. Cox J., Rubinstein M., Ross S.: *Option Pricing: A Simplified Approach*. „The Journal of Financial Economics" 1979, No. 7.
2. Jajuga K., Jajuga T.: *Inwestycje*. Wydawnictwo Naukowe PWN, Warszawa 2001.
3. Łachwa A.: *Rozmyty świat zbiorów, liczb, relacji, faktów, reguł i decyzji*. Exit, Warszawa 2001.
4. Piegat A.: *Modelowanie i sterowanie rozmyte*. EXIT, Warszawa 2003.
5. Przybycin Z.: *Zastosowanie logiki rozmytej w ekonomii – wybrane modele decyzyjne*. Akademia Ekonomiczna, Katowice 2009.
6. Weron A., Weron R.: *Inżynieria finansowa*. WNT, Warszawa 1998.

THE BINOMIAL OPTION PRICING MODEL IN CASE OF A FUZZY INFORMATION

Summary

In the article presents the modification of the binomial model of European option pricing. Adopted namely, that it is appropriate to the weakening of the assumptions about the mechanism of price formation on the underlying instrument. In the proposed option pricing model, mechanism of option pricing formation is described in terms of fuzzy numbers. In the article, also posted selected messages from the scope of fuzzy numbers and options.

Adam R. Szromek*

POMIAR FUNKCJI TURYSTYCZNEJ OBSZARÓW ZA POMOCĄ WSKAŹNIKÓW FUNKCJI TURYSTYCZNEJ NA PRZYKŁADZIE OBSZARÓW PAŃSTW EUROPEJSKICH

Wprowadzenie

Turystyka jest zjawiskiem, którego bezpośredni pomiar jest niemożliwy, tak jak niemożliwe jest zmierzenie wielkości popytu realizowanego przez turystów. Złożoność tego zjawiska wynika z szerokiego zakresu działalności i aktywności turystycznej, gdyż o ile możliwy jest pomiar liczby turystów, osobonoclegów czy nawet dochodów uzyskanych z działalności turystycznej, o tyle nie sposób zmierzyć, jaka część obrotów handlowych wynika ze sprzedaży dóbr i usług konsumowanych przez turystów, a jaka miałaby miejsce, gdyby ruch turystyczny nie występował. Niezwykle trudny wydaje się być również pomiar eksploatacji zasobów infrastrukturalnych w miejscowości turystycznej, spowodowany pobytem lub nawet przejazdem turysty przez tę miejscowość. Nie bez znaczenia w tym względzie są podejmowane w ostatniej dekadzie próby wprowadzenia tzw. rachunku satelitarnego. Wciąż jest to jednak szacowanie, a nie obliczanie wielkości popytu.

Próby szacowania wielkości popytu turystycznego były podejmowane już pół wieku temu, kiedy to zainicjowano użycie miar o charakterze symptomatycznym. Do dziś stosuje się takie wielkości, jak liczba turystów, osobonoclegów czy miejsc noclegowych. Nie są to jednak miary możliwe do wykorzystania w sposób bezpośredni, inaczej czołówkę wszelkich klasyfikacji zajmowałyby obszary, gdzie występuje najwięcej turystów bez względu na wielkość miejscowości. Potrzeba odniesienia poszczególnych zmiennych związanych z turystyką

* Publikacja powstała w ramach projektu badań własnych nr 3709/B/H03/2011/40 pt. *Wskaźniki funkcji turystycznej obszaru recepcji turystycznej*. Projekt został sfinansowany ze środków Narodowego Centrum Nauki na podstawie decyzji Ministra Nauki i Szkolnictwa Wyższego RP z dnia 4 lutego 2011 roku.

do zmiennych będących miarami odniesienia, tj. powierzchnia obszaru lub liczba mieszkańców tego obszaru, skłoniła badaczy do opracowania wskaźników funkcji turystycznej.

Niniejszy artykuł prezentuje najpopularniejsze z nich i uzupełnia je nowoczesnymi miarami syntetycznymi. Przedmiotem zastosowania omawianych wskaźników jest pomiar funkcji turystycznej w 31 państwach europejskich oraz dokonanie ich klasyfikacji zgodnie z wartościami zastosowanych miar. Nietypowy obszar badań ujmujący zagadnienie w skali kraju, a nie w ujęciu lokalnym pozwala na znacznie bardziej wyraźne ukazanie słabości poszczególnych miar ujmujących opisywane zagadnienie.

1. Funkcja turystyczna i jej pomiar

Funkcja turystyczna (wypoczynkowa) obszaru to szeroko rozumiana działalność społeczno-ekonomiczna pełniona przez określony obszar (region lub miejscowość) i jego mieszkańców w celu przywrócenia osobom przybyłym utraconych w wyniku zmęczenia normalnych zdolności psychofizycznych ich organizmu z wykorzystaniem naturalnych właściwości środowiska geograficznego [5]. A. Kowalczyk [4] uzupełnia powyższą definicję i stwierdza, że funkcja turystyczna to działalność społeczno-ekonomiczna, która jest skierowana na obsługę turystów i którą miejscowość lub obszar spełnia w systemie gospodarki narodowej.

Funkcję turystyczną obszaru analizuje się w wielu aspektach prowadzonej działalności. Literatura zawiera przykłady miar diagnozujących tę rolę obszaru ze względu na intensywność ruchu turystycznego oraz zagospodarowanie turystyczne, a czasem również za pomocą syntezy obu zagadnień.

Analiza rozwoju funkcji turystycznej zwykle dotyczy ograniczonego obszaru będącego odpowiednio zdefiniowanym terytorium. Zwykle jest to obszar określony podziałem administracyjnym (gmina, powiat, województwo). Nie spotyka się natomiast większych (niepodzielnych) obszarów badań, jak państwo czy grupa państw. Głównym tego powodem jest mała istotność takich porównań dla gospodarki pojedynczego państwa. Jednak w sytuacji gdy porównanie takie stanowi element umożliwiający klasyfikację państw ze względu na stopień rozwoju funkcji turystycznej, a zatem jest narzędziem wskazania krajów najbardziej rozwiniętych pod względem turystycznym, wówczas informacja taka staje się cenna ze względu na możliwość pozyskania odpowiedniego wzorca rozwoju.

Pomiar rozwoju funkcji turystycznej jest dokonywany za pomocą kilku wskaźników, wśród których należy wymienić najlepiej znane tradycyjne wskaźniki funkcji turystycznej, tj.: Baretje'a-Deferta, Schneidera, Charvata czy Deferta.

Wskaźnik Baretje'a-Deferta (W_{BD}) [2] jest zaliczany do grupy mierników zagospodarowania turystycznego. Oblicza się go przez odniesienie wielkości określającej liczbę miejsc noclegowych do liczby rezydentów obszaru (1):

$$W_{BD} = \frac{liczba\ turystycznych\ miejsc\ noclegowych}{liczba\ stałych\ mieszkańców\ obszaru} \cdot 100 \quad (1)$$

Natomiast wskaźnik Deferta (W_D) to wskaźnik dostarczający informacji o liczbie turystów przypadających na 1 km^2 obszaru turystycznego (2), który pozwala ocenić zatłoczenie (gęstość populacji) badanego obszaru [3]:

$$W_D = \frac{liczba\ turystów\ korzystających\ z\ noclegów}{powierzchnia\ obszaru\ w\ km^2} \quad (2)$$

Wskaźnik Schneidera (W_{Sch}) jest miernikiem intensywności ruchu turystycznego, gdyż wyraża liczbę turystów korzystających z noclegów przypadających na 100 stałych mieszkańców obszaru. Jego postać jest następująca (3):

$$W_{Sch} = \frac{liczba\ turystów\ korzystających\ z\ noclegów}{liczba\ stałych\ mieszkańców\ obszaru} \cdot 100 \quad (3)$$

Z kolei wskaźnik Charvata (W_{Ch}) jest miarą określającą liczbę udzielonych noclegów przypadających na 100 mieszkańców obszaru. Pozwala zatem ocenić intensywność ruchu turystycznego, jaka występuje na badanym obszarze (4):

$$W_{Ch} = \frac{liczba\ udzielonych\ noclegów}{liczba\ stałych\ mieszkańców\ obszaru} \cdot 100 \quad (4)$$

Inną często używaną miarą rozwoju funkcji turystycznej jest wskaźnik gęstości bazy noclegowej (W_{GBN}). Pozwala on określić zagęszczenie miejsc noclegowych na badanym obszarze (5) [1; 8]:

$$W_{GBN} = \frac{liczba\ turystycznych\ miejsc\ noclegowych}{powierzchnia\ obszaru\ w\ km^2} \quad (5)$$

Warto zauważyć, że wymienione wskaźniki bazują zaledwie na kilku zmiennych powtarzających się w różnych konfiguracjach i z różnymi miarami odniesienia, ale literatura wymienia jeszcze inne miary rozwoju obszaru turystycznego. Jedną z najbardziej interesujących miar jest wskaźnik J. Warszyńskiej [7], która zauważa, iż poza poziomem rozwoju obszaru turystycznego warto określić również stopień jej zdefiniowania. Chodzi o sytuację, w której określony obszar ma niejednoznaczne oceny (w różnych aspektach rozwoju funkcji turystycznej), tzn. jeden ze wskaźników wskazuje na wysoko rozwiniętą rolę turystyki, a inny na słaby poziom tej branży. J. Warszyńska proponuje wówczas określić zgodność wskaźników i włączyć tę informację w ocenę obszaru.

Wykorzystanie tej cechy wymagało opracowania określonego sposobu kwantyfikowania zgodności wskaźników. Zgodnie z koncepcją zaproponowaną przez autora [6] jest możliwe utworzenie wskaźnika syntetycznego, którego konstrukcja wynika właśnie z syntezy oceny poziomu rozwoju funkcji turystycznej oraz stopnia zdefiniowania tej funkcji. Dzięki temu powstał dwuwymiarowy wskaźnik funkcji turystycznej.

Tabela 1 prezentuje kryteria szacowania wartości wskaźnika rozwoju funkcji turystycznej (W_{RFT}) będącego pierwszym (cząstkowym) wskaźnikiem funkcji turystycznej. Powstaje on poprzez uśrednienie uzyskanych wyników (P_{RFT}) podanych w skali od 0 do 1 (gdzie 1 oznacza bardzo dobrze rozwiniętą funkcję turystyczną). Podstawą pomiaru jest tu grupa wymienionych wcześniej czterech wskaźników funkcji turystycznej, opisanych wzorami: (1) (2) (3) i (5). Są to zatem dwa wskaźniki intensywności ruchu turystycznego oraz dwa wskaźniki zagospodarowania turystycznego.

Tabela 1

Kryteria klasyfikacji funkcji turystycznej obszarów

Pozycja określająca stopień rozwoju funkcji turystycznej – stymulanta (P_{RFT} (P_X))	**Poziom zagospodarowania turystycznego**		**Intensywność ruchu turystycznego**		**wskaźnik rozwoju funkcji turystycznej (W_{RFT})**
	wskaźnik Baretje'a-Deferta (W_{BD})	**wskaźnik gęstości bazy noclegowej (W_{GBN})**	**wskaźnik Schneidera (W_{Sch})**	**wskaźnik Deferta (W_D)**	
	miejsca noclegowe		**liczba turystów**		
	na 100 mieszkańców	**na 1 km^2 powierzchni**	**na 100 mieszkańców**	**na 1 km^2 powierzchni**	***$(P_{BD}+P_{GBN}+P_{Sch}+P_D)/16$***
0	0,00-0,78	0,00-0,78	0,0-7,8	0,0-15,6	0,00-0,10
1	0,78-6,25	0,78-6,25	7,8-62,5	15,6-125	0,10-0,30
2	6,25-25,00	6,25-25,00	62,5-250	125-500	0,30-0,60
3	25,00-50,00	25,00-50,00	250-500	500-1000	0,60-0,80
4	50,00 i więcej	50,00 i więcej	500 i więcej	1000 i więcej	0,80-1,00

Źródło: Opracowanie własne na podstawie [6; 7].

Z kolei miarą zdefiniowania funkcji turystycznej jest jednoznaczność wskazań mierników rozwoju funkcji turystycznej (D_{RFT}). Również ten aspekt sprowadzono do czterostopniowej skali ocen, których kwantyfikacja opiera się na porównaniu otrzymanego wyniku klasyfikacji (dokonanej za pomocą czterech zastosowanych wskaźników) z opracowanym wzorcem kombinacji. Interpretacji stopnia zdefiniowania funkcji dokonuje się za pomocą tabeli 2, która prezentuje

wszystkie możliwe kombinacje wskazań wraz z ich ocenami (gdzie *x* oznacza powtarzającą się pozycję określającą stopień rozwoju funkcji turystycznej obszaru, natomiast *o* to jakakolwiek inna wartość pozycji niż *x*).

Tabela 2

Kombinacja możliwych pozycji i przypisany im stopień zdefiniowania funkcji turystycznej

Kombinacja wyniku (K_{RFT})	*xxxx*	*xxxo*	*xxox*	*xoxx*	*oxxx*	*xxoo*	*xoxo*	*oxox*	*ooxx*	*oxxo*	*xoox*	*Inne*
Stopień zdefiniowania funkcji turystycznej (D_{RFT})	3	2	2	2	2	1	1	1	1	1	1	0

Objaśnienia symboli: *x* – powtarzający się stopień rozwoju, *o* – inny stopień rozwoju niż *x*.

Wynik iloczynu parametrów (P_{RFT} i D_{RFT}) pozwala na uzyskanie właściwej miary syntetycznej oceniającej rozwój funkcji turystycznej w dwóch wymiarach, czyli dwuwymiarowego wskaźnika funkcji turystycznej. Miara ta jest interpretowana poprzez czterostopniową skalę rozwoju wskaźnika $W_{W\text{-}Sz}$, gdzie jego wynik od 0 do 3 wskazuje na brak rozwiniętej funkcji turystycznej badanego obszaru, wynik z przedziału od 4 do 6 umownie wskazuje na pierwszy stopień rozwoju funkcji turystycznej, wynik między 7 a 9 to drugi stopień rozwoju funkcji turystycznej, natomiast wynik równy 10 lub więcej to najwyższy (trzeci) stopień rozwoju funkcji turystycznej.

Ze względu na liczne przypadki źle zdefiniowanej funkcji turystycznej opracowano również miary syntetyczne określające rozwój turystyczny obszaru w obu analizowanych aspektach jednocześnie [6]. Jedną z takich miar jest logistyczny wskaźnik funkcji turystycznej $W_{L\beta}$ w postaci:

$$W_{L\beta} = \frac{U_n^{1,47} \cdot e^{-21,18+0,11\frac{T}{p\cdot\sqrt{Msc}}+0,71\left(\frac{\log U_n}{R}\right)^3}}{1+U_n^{1,47} \cdot e^{-21,18+0,11\frac{T}{p\cdot\sqrt{Msc}}+0,71\left(\frac{\log U_n}{R}\right)^3}} \tag{6}$$

gdzie:
Un – liczba udzielonych noclegów,
Msc – liczba turystycznych miejsc noclegowych,
R – liczba stałych mieszkańców obszaru,
T – liczba turystów korzystających z noclegów,
P – powierzchnia badanego obszaru w km^2.

Powstanie wskaźników syntetycznych wynikało z potrzeby ogólnego, ale zarazem możliwie pełnego ujęcia problemu oceny rozwoju turystycznego obszaru. O ile takie wskaźniki, jak wskaźnik Baretje'a-Deferta czy Charvata, oceniają odmienne aspekty tego złożonego zjawiska, o tyle miary syntetyczne umożliwiają uchwycenie różnych aspektów w jednej ocenie. Taką rolę pełni właśnie logistyczny wskaźnik funkcji turystycznej, a także dwuwymiarowy wskaźnik funkcji turystycznej.

2. Analiza rozwoju turystycznego państw europejskich

Analizy rozwoju funkcji turystycznej obszarów (w ujęciu obszaru państwa) dokonano na podstawie danych EUROSTAT. Bazę danych utworzono z takich zmiennych, jak:

- liczba turystów rejestrowanych w każdym kraju,
- liczba osobonoclegów zrealizowanych w ramach pobytów turystycznych w hotelach w każdym z krajów,
- ilość miejsc noclegowych oferowanych turystom w całym kraju,
- powierzchnia kraju,
- liczba ludności (rezydentów).

Niestety dane za 2010 rok dotyczące powyższych wielkości są niekompletne. Braki danych dotyczące niektórych państw wymusiły dokonanie dodatkowych przekształceń umożliwiających klasyfikację państw europejskich. Pełną bazę danych uzyskano poprzez uśrednienie danych z szeregu czasowego z okresu 5 ostatnich lat (tj. 2006-2010). Tym samym uzyskane wyniki dotyczą poziomu rozwoju funkcji turystycznej w ostatnich 5 latach.

Uchwycenie poszczególnych aspektów rozwoju funkcji turystycznej państw wymagało zastosowania zróżnicowanych wskaźników, wśród których znalazł się wskaźnik Charvata, wskaźnik Deferta, wskaźnik Schneidera, wskaźnik Baretje'a-Deferta oraz wskaźnik gęstości bazy noclegowej, a także dwa również omówione wskaźniki, czyli dwuwymiarowa miara rozwoju turystycznego J. Warszyńskiej [7] (z poprawką interpretacyjną autora [6]) $W_{W\text{-}Sz}$ oraz logistyczny wskaźnik funkcji turystycznej $W_{L\beta}$ [6].

Wartości poszczególnych wskaźników funkcji turystycznej odzwierciedlają stopień rozwoju turystyki na obszarach państw ogółem, a zatem uwzględniają całe terytorium państwa, a nie wyłącznie obszary turystyczne. Stąd też uzyskane wyniki są znacząco niższe od uśrednionych wartości wskaźników dla obszarów turystycznych, gdyż ujmują również obszary, na których nie występuje zjawisko turystyki. Można to zaobserwować w tabelach 3-5, porównując poszczególne wyniki dla Polski z wynikami uzyskanymi w innych badaniach autora [6], w których dokonano uśrednienia wartości wskaźników wyłącznie dla obszarów turystycznych.

Intensywność ruchu turystycznego poszczególnych państw prezentują tabele 3 i 4, w których ujęto wartości wskaźników intensywności turystycznej za pomocą wskaźników Charvata, Deferta i Schneidera.

Tabela 3

Klasyfikacja zgodnie z wartością wskaźnika Charvata i Deferta

Państwo	Wskaźnik Charvata	Państwo	Wskaźnik Deferta
Norwegia	273,63	Holandia	224,43
Austria	259,73	Belgia	137,50
Hiszpania	246,91	Niemcy	134,30
Włochy	232,92	Wielka Brytania	120,68
Szwecja	212,12	Szwajcaria	112,30
Niemcy	210,58	Luksemburg	101,08
Finlandia	207,67	Włochy	81,69
Francja	200,21	Dania	66,54
Szwajcaria	199,85	Cypr	61,51
Irlandia	190,33	Republika Czeska	61,35
Wielka Brytania	163,34	Francja	60,81
Cypr	151,73	Słowenia	50,75
Grecja	146,24	Austria	50,15
Islandia	128,64	Słowacja	48,63
Portugalia	122,90	Węgry	38,90
Dania	109,73	**P o l s k a**	35,69
Holandia	103,60	Hiszpania	32,67
Słowenia	98,99	Grecja	30,88
Republika Czeska	88,61	Portugalia	21,78
Malta	80,99	Rumunia	19,30
Estonia	78,40	Szwecja	18,27
Węgry	74,65	Litwa	13,12
Rumunia	69,36	Estonia	9,28
Słowacja	63,08	Finlandia	7,58
Bułgaria	62,94	Łotwa	5,59
Chorwacja	61,63	Bułgaria	4,05
Belgia	50,10	Irlandia	brak danych
P o l s k a	43,63	Malta	brak danych
Łotwa	36,02	Islandia	brak danych
Litwa	28,12	Norwegia	brak danych
Luksemburg	16,03	Chorwacja	brak danych

Warto zauważyć, iż miejsce Polski w tych ujęciach jest odległe, jednak fakt ten nie jest zaskoczeniem, szczególnie w porównaniu z takimi państwami, jak

Włochy czy Hiszpania, gdzie ruch turystyczny jest znacząco wyższy niż w Polsce. Dziwić może wyższa pozycja Polski niż Hiszpanii w przypadku wskaźnika Deferta. Dzięki temu zestawieniu można zauważyć słabość tej miary, gdyż (jak ukazuje tabela 3) czołówkę tej klasyfikacji stanowią państwa o niewielkiej powierzchni (Belgia, Holandia, Luksemburg, Szwajcaria itp.). Niska pozycja Hiszpanii, Portugalii i wielu innych państw wynika ze znacząco większej powierzchni i populacji mieszkańców w tych krajach. Nie zauważa się już tej tendencji w przypadku klasyfikacji dokonanej za pomocą wskaźnika Charvata (co oczywiście wynika z odmiennej konstrukcji tego wskaźnika, gdyż wykorzystuje on wielkość populacji państwa, a nie jego powierzchnię). Tam Hiszpania i Włochy są już w czołówce państw europejskich.

Podobną klasyfikację zauważa się w przypadku wskaźnika Schneidera, który również mierzy ruch turystyczny w relacji do liczby mieszkańców obszaru. W tym zestawieniu (tabela 4) czołówkę państw o największej intensywności ruchu turystycznego stanowią takie kraje, jak Szwecja czy Cypr. Ponownie wydaje się, że wpływ wielkości populacji poszczególnych państw ma ogromne znaczenie dla szacowanej oceny. Interesująca jest pozycja Niemiec we wszystkich trzech rankingach oceniających intensywność ruchu turystycznego. Państwo to znajduje się na 6, 3 i 4 miejscu w kolejnych klasyfikacjach i wyraźniej ugruntowało swoją pozycję rankingową.

Tabela 4

Klasyfikacja zgodnie z wartością wskaźnika Schneidera i Baretje'a-Deferta

Państwo	**Wskaźnik Schneidera**	**Państwo**	**Wskaźnik Baretje'a-Deferta**
1	2	3	4
Szwecja	88,07	Cypr	10,73
Cypr	71,22	Malta	9,51
Szwajcaria	59,57	Austria	6,93
Niemcy	58,63	Grecja	6,38
Holandia	56,19	Islandia	5,86
Luksemburg	52,34	Hiszpania	3,68
Dania	51,87	Włochy	3,62
Francja	51,12	Chorwacja	3,58
Słowenia	50,33	Irlandia	3,53
Austria	50,24	Szwajcaria	3,50
Finlandia	47,95	Norwegia	3,32
Wielka Brytania	47,30	Bułgaria	3,11
Republika Czeska	46,07	Luksemburg	2,95
Słowacja	43,92	Portugalia	2,55
Włochy	40,79	Republika Czeska	2,40

cd. tabeli 4

1	2	3	4
Belgia	38,69	Szwecja	2,30
Węgry	36,13	Finlandia	2,24
Grecja	36,05	Estonia	2,18
Hiszpania	35,95	Wielka Brytania	2,08
Estonia	31,31	Niemcy	2,05
P o l s k a	29,24	Francja	1,94
Litwa	25,74	Słowenia	1,92
Rumunia	21,44	Węgry	1,57
Portugalia	18,86	Dania	1,36
Łotwa	16,07	Słowacja	1,27
Bułgaria	5,94	Holandia	1,22
		Belgia	1,16
		Rumunia	1,12
		Łotwa	1,04
		Litwa	0,68
		P o l s k a	0,55

Zupełnie inną klasyfikację uzyskuje się w przypadku, gdy bada się poziom zagospodarowania turystycznego wskaźnikiem Baretje'a-Deferta. Ponownie czołowe miejsca zajmują państwa najmniejsze, a właściwie o najmniejszej populacji mieszkańców. Fakt ten potwierdza kolejny ze wskaźników (gęstości bazy noclegowej). Ponownie najmniejsze państwa znalazły się w czołówce państw najlepiej rozwiniętych turystycznie pod względem infrastruktury turystycznej (tabela 5). Nie oznacza to, że wyniki są nieprawdziwe, ale ujawnia się złożoność zjawiska, które jest mierzone w różnych aspektach turystycznych.

Uzyskane wyniki klasyfikacji poszczególnych państw są bardzo zróżnicowane, a uzyskane rozbieżności wynikające z obserwacji rynku turystycznego mogą sugerować w wielu przypadkach zaistnienie wątpliwości co do uzyskanych wyników jednowymiarowych, dlatego warte uwagi jest wykorzystanie wskaźników mierzących analizowane zjawisko w sposób syntetyczny, a zatem z jednoczesnym ujęciem intensywności ruchu turystycznego i zagospodarowania turystycznego obszarów.

Tabela 5

Klasyfikacja państw europejskich zgodnie z wartością wskaźnika gęstości bazy noclegowej oraz logistycznego wskaźnika funkcji turystycznej

Państwo	Wskaźnik gęstości bazy	Państwo	Logistyczny wskaźnik funkcji turystycznej
Malta	131,33	Niemcy	0,999
Cypr	9,27	Włochy	0,998
Włochy	7,24	Francja	0,998
Austria	6,92	Hiszpania	0,998
Szwajcaria	6,59	Wielka Brytania	0,997
Luksemburg	5,69	Austria	0,976
Grecja	5,46	Szwecja	0,972
Wielka Brytania	5,30	Holandia	0,967
Holandia	4,85	Grecja	0,963
Niemcy	4,69	**P o l s k a**	0,962
Belgia	4,12	Szwajcaria	0,961
Hiszpania	3,35	Rumunia	0,957
Republika Czeska	3,19	Portugalia	0,948
Portugalia	2,95	Finlandia	0,937
Chorwacja	2,80	Republika Czeska	0,918
Francja	2,30	Węgry	0,890
Irlandia	2,24	Dania	0,859
Bułgaria	2,12	Belgia	0,837
Słowenia	1,93	Bułgaria	0,805
Dania	1,74	Słowacja	0,721
Węgry	1,69	Słowenia	0,549
Słowacja	1,40	Cypr	0,365
Rumunia	1,00	Estonia	0,311
P o l s k a	0,67	Litwa	0,254
Estonia	0,65	Łotwa	0,225
Szwecja	0,48	Luksemburg	0,008
Norwegia	0,42		
Łotwa	0,36		
Finlandia	0,35		
Litwa	0,35		
Islandia	0,18		

Jedną z takich miar jest logistyczny wskaźnik funkcji turystycznej będący miarą opartą na kilku zmiennych będących podstawą konstrukcyjną poprzednio omawianych tradycyjnych wskaźników funkcji turystycznej. Miara ta umożliwiła uzyskanie klasyfikacji syntetycznej, w której ocenę ujmuje się w skali od 0 do 1 (będącą stymulantą).

W czołówce tej klasyfikacji znalazły się Niemcy, Włochy i Francja oraz Hiszpania (tabela 5). Można zauważyć, iż to zestawienie wydaje się być bardziej wiarygodne, niż poprzednio użyte miary cząstkowe. Interesująca jest również pozycja państw najmniejszych – stanowią one zazwyczaj ostatnie miejsca tej klasyfikacji mimo faktu, iż wskaźnik uwzględnia zarówno wielkość powierzchni państwa, jak i liczbę obywateli. Polska w tym zestawieniu zajmuje 10 pozycję rankingową, a zatem jest to wynik znacząco lepszy niż ten, jaki uzyskano w poprzednich zestawieniach.

Uwzględniając zarówno poziom funkcji turystycznej oraz jej zdefiniowanie za pomocą dwuwymiarowego wskaźnika funkcji turystycznej, uzyskano jedynie wyniki przybliżone ze względu na znacznie niższe miary cząstkowe. Analiza wymienia 5 państw, gdzie rozwój turystyki jest większy niż w pozostałych krajach (tabela 6). Są to: Belgia, Niemcy, Włochy, Cypr i Holandia. Niestety ta klasyfikacja jest niepełna ze względu na fakt, iż obliczenie oceny wymaga użycia wszystkich wymienionych wcześniej miar. Ponieważ nie zawsze możliwe było obliczenie wszystkich wskaźników, więc klasyfikacja sporządzona miarą dwuwymiarową jest niekompletna.

Tabela 6

Klasyfikacja zgodnie z wartością dwuwymiarowego wskaźnika funkcji turystycznej

Państwo	**P_{RFT} (stopień rozwoju funkcji turystycznej)**	**D_{RFT} (stopień zdefiniowania funkcji turystycznej)**	**Dwuwymiarowy wskaźnik funkcji turystycznej $W_{W\text{-}Sz}$**
1	2	3	4
Belgia	2	2	1
Niemcy	2	2	1
Włochy	2	2	1
Cypr	2	2	1
Holandia	2	2	1
Austria	2	1	0
Republika Czeska	1	3	0
Dania	1	3	0
Hiszpania	1	3	0
Luksemburg	1	3	0
Węgry	1	3	0
Portugalia	1	3	0
Rumunia	1	3	0
Słowenia	1	3	0
Słowacja	1	3	0
Szwajcaria	1	2	0
Bułgaria	1	1	0

cd. tabeli 6

1	2	3	4
Estonia	1	1	0
Łotwa	1	1	0
P o l s k a	1	1	0
Finlandia	1	1	0
Litwa	0	2	0

Tabela 6 pokazuje, iż Polska jako kraj turystyczny znajduje się na odległym miejscu. Wynika to z niskiego poziomu rozwoju funkcji turystycznej, a także ze znaczącej zmienności tego poziomu na różnych obszarach. Państwa wymienione w czołówce rankingu są dobrze rozwinięte pod względem turystycznym, a zarazem charakteryzuje je dobre zdefiniowanie funkcji turystycznej. Ugruntowana jest również pozycja Litwy jako państwa zupełnie nieposiadającego aspiracji do miana obszaru turystycznego Europy.

Podsumowanie

Podsumowując poszczególne wyniki uzyskane dla Polski, warto wspomnieć, iż nie powinny być one porównywane z wynikami uzyskanymi dla mniejszych obszarów (gmin, województw), gdyż uwzględniają obszary nieturystyczne. Badania zrealizowane przez autora w 2011 roku [6] obejmowały analizę funkcji turystycznej obszarów gmin całej Polski, co zarazem umożliwiło uzyskanie średnich wartości każdego ze wskaźników funkcji turystycznej. Przeciętna wartość wskaźnika Deferta wyniosła 161, wskaźnika Schneidera 89,5, wskaźnika Charvata 339,2, wskaźnika Bretje'a-Deferta 4,6, a wskaźnika gęstości bazy noclegowej 5,8.

Porównując poszczególne miary funkcji turystycznej zauważa się, że nie wskazują one jednoznacznych wyników, a zatem użycie miary syntetycznej wydaje się być właściwym rozwiązaniem. Wobec niekompletności wyników uzyskanych za pomocą dwuwymiarowego wskaźnika funkcji turystycznej za wysoce istotną miarę należy uznać logistyczny wskaźnik funkcji turystycznej, który uwzględnia zarówno wielkość obszaru, jak i jego zaludnienie. Jednocześnie nie pozwala on, aby zmienne będące jedynie miarami odniesienia wpływały na wynik końcowy w takim stopniu, jak w przypadku tradycyjnych wskaźników funkcji turystycznej.

Literatura

1. Baran E., Grzebyk B.: *Uwarunkowania rozwoju turystyki na obszarach Podkarpacia o niekorzystnych warunkach gospodarowania*. „ACTA SCIENTIARUM POLONORUM

Oeconomia" 9 (4), Bydgoszcz, Kraków, Lublin, Olsztyn, Poznań, Siedlce, Szczecin, Warszawa, Wrocław 2010, s. 15-24.

2. Baretje R., Defert P.: *Aspects economiques du tourisme*. Paris 1972.
3. Defert P.: *Le Taut de FonctionTouristique: mise au point et critique*. W: *Les Cahiers du Tourisme. Aux-en-provence*. Centre des Hautes Etudes Touristiques, Berne 1967, s. C-13.
4. Kowalczyk A.: *Geografia turyzmu*. Wydawnictwo Naukowe PWN, Warszawa 2000.
5. Matczak A.: *Funkcja wypoczynkowa strefy podmiejskiej Łodzi.* Zakład Geografii Miast i Turyzmu, UŁ, Łódź 1982 (maszynopis niepublikowany).
6. Szromek A.R.: *Wskaźniki funkcji turystycznej obszarów recepcji turystycznej*. Wydawnictwo Politechniki Śląskiej, Gliwice 2012 (praca w druku).
7. Warszyńska J.: *Funkcja turystyczna Karpat polskich*. Folia Geographica, Series Geographica-Oeconomica, Vol. XVIII, 1985, s. 88.
8. Warszyńska J. Jackowski A.: *Podstawy geografii turyzmu*. PWN, Warszawa 1979.

THE MEASUREMENT OF TOURISM FUNCTION OF AREAS USING INDICATORS OF THE TOURISM FUNCTION ON THE EXAMPLE OF AREAS OF EUROPEAN COUNTRIES

Summary

This paper presents the most popular indicators of the tourist function , and supplements them with modern synthetic metrics. The subject of the application of these indicators is to measure of the tourism function in 31 European countries and making their classification according to the values of measurement used. Unusual problem of engaging research area in the country and not in the local formulation allows for a much more clear showing weaknesses of individual metrics that recognize the problem described. A fragmented approach to the tourism function, prompts to use of synthetic measures which prove to be more useful in classifying particular areas.

Joanna Trzęsiok

OCENA WPŁYWU ZMIENNYCH OBJAŚNIAJĄCYCH NA ZMIENNĄ ZALEŻNĄ W METODZIE RZUTOWANIA PPR

Wprowadzenie

Metoda rzutowania PPR (projection pursuit regression), zaproponowana przez J. Friedmana i W. Stuetzle'a w 1981 roku, jest jedną z nieparametrycznych metod regresji. Przeprowadzone badania porównawcze pokazują, iż modele regresji otrzymane za jej pomocą charakteryzują się często wyższą dokładnością predykcji niż modele generowane przez inne metody, zarówno nieparametryczne, jak i klasyczne (zob. [4; 6; 7]). Jest to jednak jedna z metod, które często określa się mianem „czarnej skrzynki". Wyniki, które otrzymuje się przy jej użyciu, nie są zazwyczaj interpretowalne, dlatego tak ważne są wszelkie próby uzyskania dodatkowych informacji z otrzymanego modelu.

W artykule przedstawiono dwie procedury: eliminacji i dołączania zmiennych, które redukują złożoność modelu otrzymanego metodą PPR, pozwalają na wyodrębnienie zmiennych, które mają największy wpływ na zmienną zależną, jak również powiększają zasób informacji uzyskanych ze zbudowanego modelu.

Celem artykułu, jak również wspomnianych procedur eliminacji i dołączania zmiennych będzie zbudowanie rankingu zmiennych objaśniających pod względem ich siły wpływu na zmienną Y.

1. Metoda rzutowania PPR

Celem metody rzutowania jest transformacja danych z przestrzeni wielowymiarowej w przestrzeń o niższym wymiarze, w której łatwiej jest badaczowi zaobserwować pewne własności analizowanego zbioru obserwacji. Transformacja ta odbywa się poprzez zrzutowanie wektora zmiennych objaśniających $\mathbf{X}$ w kierunkach $\boldsymbol{\alpha}_k$. Uzyskuje się w ten sposób nowe zmienne:

$$Z_k = \boldsymbol{\alpha}_k^T \cdot \mathbf{X}, \text{ dla } k = 1,...,K, \tag{1}$$

gdzie $\boldsymbol{\alpha}_k \in \mathbf{R}^n$ są unormowanymi wektorami nazywanymi kierunkami rzutowania.

Model regresyjny, zbudowany za pomocą metody rzutowania, można przedstawić w postaci addytywnej:

$$Y = f(\mathbf{X}) = \alpha_0 + \sum_{k=1}^{K} \beta_k g_k\left(\boldsymbol{\alpha}_k^T \cdot \mathbf{X}\right). \tag{2}$$

Funkcje składowe modelu g_k (dla $k = 1,\ldots,K$) to funkcje jednej zmiennej, nazywane funkcjami grzbietowymi, o parametrach β_k. Estymatory tych parametrów, a także kierunków rzutowania $\boldsymbol{\alpha}_k$ otrzymuje się w kolejnych krokach algorytmu poprzez minimalizację błędu empirycznego (empirical risk):

$$R_{emp}(\boldsymbol{\alpha},\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - f(\mathbf{x}_i)\right)^2, \tag{3}$$

gdzie $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_K)$ oraz $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_K)$.

Błąd empiryczny przedstawiony we wzorze (3) można przekształcić do następującej postaci (zob. [5]):

$$R_{emp}(\boldsymbol{\alpha},\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\left(r_i - \beta_k g_k\left(\boldsymbol{\alpha}_k^T \cdot \mathbf{x}_i\right)\right)^2, \tag{4}$$

gdzie:

$$r_i = y_i - \alpha_0 - \sum_{l \neq k} \beta_l g_l\left(\boldsymbol{\alpha}_l^T \cdot \mathbf{x}_i\right) \text{ dla } i = 1,...,n. \tag{5}$$

Otrzymano w ten sposób dekompozycję błędu empirycznego na dwa składniki: resztę częściową r_i opisującą zmienność, która nie została wyjaśniona przez funkcje składowe g_l (dla $l \neq k$), oraz funkcję g_k.

Algorytm minimalizacji błędu R_{emp} jest pewnym uogólnieniem metody wykorzystującej sprzężenie zwrotne (backfitting algorithm) i składa się z następujących kroków [1, s. 255-259; 5]:

1. Ustal początkowe wartości współrzędnych wektorów $\boldsymbol{\alpha}_k$ oraz β_k (dla $k = 1,...,K$) tak, aby:

$$\beta_k g_k\left(\boldsymbol{\alpha}_k^T \cdot \mathbf{x}_i\right) \equiv 0, \text{ dla } i = 1,...,n. \tag{6}$$

Przyjmij:

$$\alpha_0 = \frac{1}{n}\sum_{i=1}^{n} y_i \,. \tag{7}$$

2. Dla każdego $k = 1,...,K$ wykonaj następujące kroki:
   a) Oblicz reszty częściowe:

$$r_i = y_i - \alpha_0 - \sum_{l \neq k} \beta_l g_l\left(\boldsymbol{\alpha}_l^T \cdot \mathbf{x}_i\right), \text{ dla } i = 1,...,n \,. \tag{8}$$

   b) Wykonaj rzutowanie, aż do osiągnięcia zbieżności:
      – ustal kierunek rzutowania $\boldsymbol{\alpha}_k$ i znajdź parametr β_k minimalizujący wyrażenie:

$$\mathrm{R}_{\mathrm{emp}}(\alpha,\beta) = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\mathrm{r}_{\mathrm{i}} - \beta_{\mathrm{k}}\mathrm{g}_{\mathrm{k}}\left(\alpha_{\mathrm{k}}^{\mathrm{T}} \cdot \mathrm{x}_{\mathrm{i}}\right)\right)^2, \tag{9}$$

      – zmień współrzędne wektora $\boldsymbol{\alpha}_k$ w kierunku wyznaczonym przez wyrażenie:

$$\boldsymbol{\alpha}_k \leftarrow \boldsymbol{\alpha}_k - \gamma \cdot \nabla R_{emp}(\boldsymbol{\alpha}_k) \,, \tag{10}$$

      gdzie $\gamma > 0$.
3. Zakończ wykonywanie algorytmu, gdy jest spełnione ustalone wcześniej kryterium stopu lub gdy wartość funkcji (3) nie zmienia się znacząco.

Najmniej ugruntowanym elementem metody rzutowania jest wybór liczby funkcji składowych K w modelu regresyjnym (2). Wartość tego parametru jest zazwyczaj podawana przez użytkownika.

Poprawę jakości budowanego modelu można uzyskać wykorzystując algorytm SMART (smooth multiply additive regression), w którym badacz podaje dwie wartości parametru K:

– K_{pocz} – początkową (maksymalną) liczbę funkcji składowych,
– K_{konc} – liczbę funkcji użytych w końcowym modelu.

Stworzony zostaje model złożony z K_{pocz} funkcji g_k, który z wykorzystaniem przyjętego kryterium zostaje stopniowo przycinany, aż do uzyskania funkcji f, która jest sumą K_{konc} składowych g_k.

2. Procedura doboru zmiennych objaśniających do modelu zbudowanego metodą rzutowania PPR

Metoda rzutowania, jako nieparametryczna metoda regresji, nie zakłada znajomości rozkładu składnika losowego w modelu czy analitycznych postaci związków między zmiennymi. Jest narzędziem, które nie wymaga spełnienia wielu restrykcyjnych założeń, przez co znacząco został zwiększony jej obszar zastosowań. W praktyce często stosuję się tę metodę do analizy zbiorów danych charakteryzowanych przez dużą liczbę zmiennych. Otrzymuje się wtedy złożony model, którego współczynniki nie są interpretowalne.

Zastosowanie procedury doboru zmiennych objaśniających do modelu pozwala na istotną redukcję liczby zmiennych, a co za tym idzie – złożoności modelu. Okazuje się również, że procedura ta pozwala na poprawę dokładności predykcji, a także stworzenie rankingu zmiennych pod względem ich siły wpływu na zmienną zależną. W ten sposób badacz, czy decydent, otrzymuje prostszy model, dający mniejsze błędy prognoz, jak również dodatkową informację o tym, które zmienne są najbardziej istotne dla tego modelu.

W artykule przedstawiono dwa warianty procedury doboru zmiennych objaśniających do modelu regresyjnego: eliminację zmiennych oraz dołączanie zmiennych.

2.1. Procedura eliminacji zmiennych z modelu

Procedura eliminacji zmiennych opiera się na strategii wspinaczki. W pierwszym etapie tej procedury zostaje zbudowany model na oryginalnym zbiorze wszystkich zmiennych. W każdym kolejnym kroku zostaje usunięta jedna zmienna według ustalonego a priori kryterium i jest budowany model na pomniejszonym zbiorze zmiennych. Wykorzystywanym kryterium jest w tym przypadku minimalny błąd średniokwadratowy liczony metodą sprawdzania krzyżowego. W ten sposób kolejno są eliminowane zmienne, które mają najmniejszy wpływ na zmienną zależną. Procedura jest powtarzana tak długo, aż w zbiorze zostanie tylko jedna zmienna. Ta właśnie zmienna ma najsilniejszy wpływ na zmienną Y.

Procedurę eliminacji zmiennych z modelu można przedstawić w następujących krokach:

1. Za pomocą metody rzutowania PPR zbuduj model regresyjny f_0, wykorzystując kompletny zbiór zmiennych objaśniających:

$$V_0 = \{X_1, X_2, \ldots, X_m\}.$$

2. Dla $j = 1, \ldots, m-1$ wykonaj następujące kroki:

a) Ze zbioru zmiennych objaśniających V_{j-1} usuń tymczasowo jedną zmienną, wykonując tę czynność kolejno dla każdej ze zmiennych, i zbuduj $(m-j+1)$ modeli regresyjnych za pomocą metody PPR.
b) Dla wszystkich zbudowanych w poprzednim kroku modeli oblicz, metodą sprawdzania krzyżowego z podziałem zbioru danych na pięć części, błąd średniokwadratowy.
c) Ostatecznie w kroku j wyeliminuj zmienną, której usunięcie w najmniejszym stopniu zmieniło dokładność predykcji modelu, a więc tą, dla której obliczony błąd średniokwadratowy jest najmniejszy. Zredukowany zbiór zmiennych oznacz przez V_j, natomiast uzyskany najmniejszy błąd średniokwadratowy zapamiętaj jako MSE_j.
d) Przyjmij jako model f_j ten model regresyjny, który był zbudowany na zbiorze zmiennych oznaczonym przez V_j i któremu odpowiada błąd średniokwadratowy MSE_j.

3. Z otrzymanego ciągu modeli regresyjnych $\{f_j\}_{j=0,\ldots,m-1}$ (z malejącą liczbą zmiennych) wybierz ten model, dla którego błąd średniokwadratowy MSE_j jest najmniejszy. Jest to model końcowy zbudowany za pomocą metody rzutowania PPR z wykorzystaniem procedury eliminacji zmiennych.

W każdym kroku tej procedury zostaje wyeliminowana jedna zmienna, ta, która ma najmniejszy wpływ na zmienną zależną. Otrzymuje się zatem ranking zmiennych pod względem ich siły wpływu na zmienną Y, gdzie najbardziej istotna jest zmienna, która pozostaje na końcu w zbiorze zmiennych.

Obliczany na każdym etapie błąd średniokwadratowy pozwala na wybranie takiego modelu, któremu jest przyporządkowany najmniejszy MSE. Zmienne wykorzystane do budowy tego modelu to zmienne, które mają istotny wpływ na zmienną zależną. Pozostałe to zmienne redundantne.

W wyniku zastosowania procedury eliminacji otrzymuje się model, który jest rozwiązaniem optymalnym jedynie w sensie lokalnym. Zaletą tego podejścia jest jednak stosunkowo niska złożoność algorytmu.

2.2. Przykład ilustrujący procedurę eliminacji zmiennych z modelu

Przedstawiona procedura eliminacji zmiennych z modelu, zbudowanego metodą rzutowania, zostanie przedstawiona na przykładzie zbioru danych *Boston*. Obserwacje przedstawione w tym zbiorze zostały zebrane i opublikowane w 1978 roku przez Harrisona oraz Rubinfelda, badaczy, którzy zajmowali się wykrywaniem zależności pomiędzy cenami nieruchomości w Bostonie a jako-

ścią życia. Jest to zbiór szeroko znany i wykorzystywany do sprawdzania jakości modeli regresyjnych. Zgromadzone dane są charakteryzowane przez trzynaście zmiennych objaśniających:

crim – wskaźnik przestępstw,
zn – frakcja obszarów zaludnionych przekraczających 25 000 stóp kwadratowych,
indus – wskaźnik industrializacji,
chas – zmienna zero-jedynkowa wskazująca, czy teren znajduje się w pobliżu rzeki Charles,
nox – koncentracja tlenku azotu,
rm – średnia liczba pokoi,
age – procent budynków sprzed 1940 roku,
dis – ważona odległość do pięciu skupisk miejsc zatrudnienia w Bostonie,
rad – dostęp do autostrady,
tax – wysokość płaconych podatków,
ptratio – liczba uczniów na jednego nauczyciela,
black – procent ludności afroamerykańskiej,
lstat – procent ludności o niskim statusie społecznym.

Zmienną zależną jest Y = medv, czyli mediana wartości domu w tys. dolarów. Zbiór *Boston* składa się z 506 obserwacji.

Wyniki uzyskane poprzez zastosowanie procedury eliminacji przedstawiono w tabeli 1.

Tabela 1

Wyniki działania procedury eliminacji zmiennych

Etap	Wyeliminowana zmienna	Numery zmiennych usuniętych z modelu	*MSE*
0	–	–	14,089
1	crim	1	11,964
2	rad	1 9	14,350
3	chas	1 9 4	**11,331**
4	age	1 9 4 7	12,435
5	zn	1 9 4 7 2	12,338
6	indus	1 9 4 7 2 3	12,042
7	dis	1 9 4 7 2 3 8	14,055
8	black	1 9 4 7 2 3 8 12	14,992
9	tax	1 9 4 7 2 3 8 12 10	14,995
10	ptratio	1 9 4 7 2 3 8 12 10 11	17,814
11	nox	1 9 4 7 2 3 8 12 10 11 5	19,962
12	rm	1 9 4 7 2 3 8 12 10 11 5 6	27,242
13	lstat		

Błąd średniokwadratowy osiąga najmniejszą wartość, równą 11,331, dla modelu, z którego wyeliminowano zmienne: crim, rad, chas. Są to zmienne redundantne. Wprowadzenie ich do modelu powoduje zwiększenie wartości MSE oraz złożoności modelu. Pozostałe dziesięć zmiennych ma istotny wpływ na zmienną zależną i postać modelu.

Największy wpływ na medv ma zmienna, którą otrzymano w ostatnim, 13. kroku, natomiast najmniejsze znaczenie ma zmienna wyeliminowana w pierwszym etapie. Otrzymane wyniki pozwalają na stworzenie rankingu zmiennych objaśniających pod względem siły wpływu na zmienną zależną (zob. tabela 2).

Tabela 2

Ranking zmiennych objaśniających pod względem siły wpływu na zmienną zależną uzyskany za pomocą procedury eliminacji zmiennych

Nr w rankingu	Zmienne	
1	lstat	zmienne istotne
2	rm	
3	nox	
4	ptratio	
5	tax	
6	black	
7	dis	
8	indus	
9	zn	
10	age	
11	chas	zmienne redundantne
12	rad	
13	crim	

Największe znaczenie dla zmiennej medv ma tutaj zmienna lstat, tak więc największy wpływ na medianę wartości domu ma procent ludności o niskim statusie społecznym. Kolejną ważną zmienną jest rm – średnia liczba pokoi.

2.3. Procedura dołączania zmiennych do modelu

Alternatywnym podejściem do eliminacji zmiennych z modelu regresyjnego jest procedura dołączania zmiennych do modelu. Zaczyna się w tym przypadku od modelu zbudowanego dla jednej zmiennej, by sukcesywnie dołączać do niego kolejne zmienne i na końcu otrzymać model zbudowany na kompletnym zbiorze zmiennych.

W pierwszym etapie tej procedury buduje się m modeli dla pojedynczych zmiennych (gdzie m jest liczbą zmiennych objaśniających). Wybiera się z nich

najlepszy i w każdym kolejnym etapie dołącza się do niego zmienną według ustalonego a priori kryterium, którym ponownie jest minimalny błąd średniokwadratowy.

Procedurę dołączania zmiennych do modelu można przedstawić w następujący sposób:

1. Za pomocą metody rzutowania PPR zbuduj m modeli regresyjnych dla pojedynczych zmiennych objaśniających. Dla każdego modelu oblicz błąd średniokwadratowy metodą sprawdzania krzyżowego. Model, który odpowiada najmniejszej wartości MSE, przyjmij jako model początkowy f_1, zaś ze zmiennej wykorzystanej do budowy modelu f_1 stwórz początkowy, jednoelementowy zbiór zmiennych V_1. Pozostałe zmienne niech tworzą zbiór W_{m-1}.
2. Dla $j = 2,\ldots,m$ wykonaj kroki:
   a) Do zbioru zmiennych objaśniających V_{j-1} dodaj tymczasowo jedną zmienną ze zbioru W_{m-j+1}, wykonując tę czynność kolejno dla każdej zmiennej, i zbuduj $(m-j+1)$ modeli regresyjnych za pomocą metody PPR.
   b) Dla wszystkich zbudowanych w poprzednim kroku modeli oblicz, metodą sprawdzania krzyżowego z podziałem zbioru danych na pięć części, błąd średniokwadratowy.
   c) Ostatecznie w kroku j dołącz do modelu tę zmienną, dla której obliczony błąd średniokwadratowy jest najmniejszy. Powiększony zbiór zmiennych tworzących model oznacz przez V_j, pozostałe zmienne przez W_{m-j}. Uzyskany najmniejszy błąd średniokwadratowy zapamiętaj jako MSE_j.
   d) Przyjmij jako model f_j ten model regresyjny, który był zbudowany na zbiorze zmiennych oznaczonym przez V_j i któremu odpowiada błąd średniokwadratowy MSE_j.
3. Z otrzymanego ciągu modeli regresyjnych $\{f_j\}_{j=1,\ldots,m}$ (z rosnącą liczbą zmiennych) wybierz ten model, dla którego błąd średniokwadratowy MSE_j jest najmniejszy. Jest to model końcowy zbudowany za pomocą metody rzutowania PPR z wykorzystaniem procedury dołączania zmiennych.

Podobnie jak dla procedury eliminacji, można uzyskać ranking zmiennych objaśniających pod względem ich siły wpływu na zmienną zależną. Przy czym najistotniejsza tym razem jest zmienna otrzymana w pierwszym kroku procedury, natomiast najmniejsze znaczenie ma zmienna, którą dołącza się do modelu w ostatnim etapie. Zmienne, których nie wykorzystano do budowy modelu końcowego, to zmienne redundantne.

Model końcowy, otrzymany za pomocą procedury dołączania zmiennych, jest rozwiązaniem optymalnym jedynie w sensie lokalnym. Ponadto procedura dołączania zmiennych do modelu, ze względu na pierwszy etap – budowy modelu dla pojedynczej zmiennej, jest uważana za mniej stabilną niż metoda eliminacji zmiennych. Z tego też powodu jest ona rzadziej wykorzystywana w praktyce.

2.4. Przykład ilustrujący procedurę dołączania zmiennych do modelu

Ponownie, w celu ilustracji procedury dołączania zmiennych, wykorzystano zbiór danych *Boston*. Uzyskane wyniki przedstawiono w tabeli 3.

Tabela 3

Wyniki działania procedury dołączania zmiennych

Etap	Dołączona zmienna	Numery zmiennych wykorzystanych do budowy modelu	*MSE*
1	lstat	13	27,242
2	rm	13 6	19,962
3	tax	13 6 10	15,785
4	nox	13 6 10 5	14,316
5	black	13 6 10 5 12	14,134
6	age	13 6 10 5 12 7	14,135
7	dis	13 6 10 5 12 7 8	14,134
8	rad	13 6 10 5 12 7 8 9	14,134
9	ptratio	13 6 10 5 12 7 8 9 11	**14,133**
10	indus	13 6 10 5 12 7 8 9 11 3	15,643
11	chas	13 6 10 5 12 7 8 9 11 3 4	15,643
12	crim	13 6 10 5 12 7 8 9 11 3 4 1	16,298
13	zn	13 6 10 5 12 7 8 9 11 3 4 1 2	16,298

Największy wpływ na medianę wartości domu, tak samo jak poprzednio, mają zmienne: lstat oraz rm, które zostały dołączone do modelu w pierwszym i drugim kroku algorytmu. Kolejne zmienne w coraz mniejszym stopniu wpływają na zmienną zależną. Ranking wszystkich zmiennych przedstawiono w tabeli 4.

Model końcowy, w tym przypadku, to model, dla którego błąd średniokwadratowy jest równy 14,133. Do budowy tego modelu wykorzystano dziewięć zmiennych mających istotny wpływ na zmienną medv. Pozostałe zmienne: indus, chas, crim i zn są, w tym przykładzie, zmiennymi redundantnymi.

Tabela 4

Ranking zmiennych objaśniających pod względem siły wpływu na zmienną zależną uzyskany za pomocą procedury dołączania zmiennych

Nr w rankingu	Zmienne	
1	lstat	zmienne istotne
2	rm	
3	tax	
4	nox	
5	black	
6	age	
7	dis	
8	rad	
9	ptratio	
10	indus	zmienne redundantne
11	chas	
12	crim	
13	zn	

Wartość współczynnika Spearmana – zgodności uzyskanych rankingów wynosi:

$$r_S = 0{,}833 .$$

Podsumowanie

W artykule przedstawiono dwie metody doboru zmiennych objaśniających do modelu regresyjnego: eliminację oraz dołączanie zmiennych. Pomimo wbudowanego w algorytmie metody PPR mechanizmu selekcji zmiennych opartego na rzutowaniu zastosowanie omawianych metod doboru zmiennych doprowadziło do poprawy dokładności predykcji modelu. Wykorzystanie mniejszej liczby zmiennych dało w konsekwencji mniej skomplikowany model końcowy.

Systematyczna eliminacja lub dołączanie zmiennych pozwoliły na zbudowanie rankingu zmiennych objaśniających pod względem: ich siły wpływu na zmienną zależną oraz zdolności poprawiania jakości modelu PPR. W tym przypadku można także oddzielić zmienne istotne od zmiennych redundantnych. Otrzymany ranking jest również dodatkową, ważną informacją dla badacza czy decydenta posługującego się w analizie regresji metodą rzutowania PPR.

Literatura

1. Cherkassky V., Mulier F.: *Learning from Data – Concepts, Theory, and Methods*. Wiley, New York 1998.
2. Friedman J.H., Stuetzle W.: *Projection Pursuit Regression*. „Journal of the American Statistical Association" 1981, No. 76, s. 817-823.
3. Harrison D., Rubinfeld D.L.: *Hedonic Prices and the Demand for Clean Air*. „Journal of Environmental Economics and Management" 1978, No. 5, s. 81-102.
4. Meyer D., Leisch F., Hornik K.: *Benchmarking Support Vector Machines*. Report No. 78, Vienna University of Economics and Business Administration, 2002, http://www.wuwien.ac.at/am/Download/ report78.pdf.
5. Trzęsiok J.: *Metoda rzutowania w budowie modelu regresyjnego*. W: *Postępy ekonometrii*. Red. A.S. Barczak. Wydawnictwo Akademii Ekonomicznej, Katowice 2004, s. 121-130.
6. Trzęsiok J.: *Analiza wybranych własności metody MART*. W: *Taksonomia 13. Klasyfikacja i analiza danych*. Red. K. Jajuga, M. Walesiak. Prace Naukowe Akademii Ekonomicznej, Wrocław 2006, No. 1126, s. 510-518.
7. Trzęsiok J.: *Ocena zasadności łączenia wybranych nieparametrycznych modeli regresji*. W: *Taksonomia 15. Klasyfikacja i analiza danych*. Red. K. Jajuga, M. Walesiak. Prace Naukowe Uniwersytetu Ekonomicznego, Wrocław 2008, No. 1207, s. 346-353.

DETERMINING THE INFLUENCE OF PREDICTOR VARIABLES ON THE RESPONSE VARIABLE IN THE PPR MODELS

Summary

Projection Pursuit Regression (PPR) was introduced by J. Friedman and W. Stuetzle in 1981. It is one of the nonparametric regression methods. The benchmark studies show very often the superiority of PPR models over other nonparametric or classical regression models in terms of the test error. PPR produces a "black-box" prediction machine and it suffers from the lack of interpretation. Thus it seems to be an important issue to find the method for evaluating the influence of the predictor variables on the response.

We present the procedure that might be used to examine the strength of the influence of every predictor variable on the response variable in Projection Pursuit Regression models.

Michał Trzęsiok

SYMULACYJNA OCENA JAKOŚCI ZAGREGOWANYCH MODELI ZBUDOWANYCH METODĄ WEKTORÓW NOŚNYCH

Wprowadzenie

Wśród wielu nowych metod regresji wielowymiarowej na szczególną uwagę zasługuje metoda wektorów nośnych (ang. SVM – Support Vector Machines). Choć została ona pierwotnie skonstruowana do realizowania zadań dyskryminacji [10], to podobnie jak w przypadku drzew klasyfikacyjnych i regresyjnych [4], algorytm tej metody można przeformułować tak, aby służył do budowy modeli regresji. Ponadto regresyjna metoda wektorów nośnych ma wiele pożądanych własności, które przenoszą się z metody dyskryminacyjnej, tj. nie wymaga od użytkownika sprawdzania założeń o rozkładach zmiennych diagnostycznych, nieliniowość otrzymanego modelu, relatywnie małe błędy predykcji dla obserwacji ze zbioru testowego, odporność na występowanie szumu w zbiorze uczącym, bardzo liczna przestrzeń hipotez z dużą różnorodnością przeszukiwanych, potencjalnych funkcji regresji [9].

Podobnie jak w przypadku dyskryminacji, nie sposób nie wspomnieć o tych własnościach metody wektorów nośnych, które świadczą na jej niekorzyść. Mianowicie jest jedną z metod eksploracji danych, co oznacza, że jej działanie jest w dużym stopniu zautomatyzowane, podobne do „czarnej skrzynki", wyniki zaś w niewielkim stopniu poddają się interpretacji. Podsumowując, otrzymywany model, jako narzędzie służące do predykcji, charakteryzuje się dużą dokładnością, lecz znajomość analitycznej postaci modelu w niewielkim stopniu pozwala badaczowi na poznanie i zrozumienie istoty relacji między zmiennymi uwzględnionymi w modelu.

Metoda wektorów nośnych jest obecnie uważana, obok zagregowanych drzew klasyfikacyjnych, za najdokładniejsze narzędzie w dyskryminacji. W dalszej części artykułu zostanie przedstawiona metoda wektorów nośnych w regre-

sji (w możliwie krótki i zrozumiały sposób) oraz próba empirycznego zweryfikowania, na ile metoda wektorów nośnych jest konkurencyjna, pod względem dokładności predykcji, wobec innych znanych metod regresji, takich jak klasyczna regresja liniowa oraz wybrane metody nieparametryczne: metoda rzutowania, sieci neuronowe, drzewa regresyjne oraz modele zagregowane łączące wiele modeli drzew. Ponadto zostanie uwzględniona metoda wektorów nośnych w ujęciu wielomodelowym. W podejściu tym, zgodnie z koncepcją Breimana [1] łączenia równoległego modeli przez rozszerzanie, jest tworzonych wiele składowych modeli regresji, zaś predykcja dla nowych obserwacji ze zbioru rozpoznawanego jest wyznaczana poprzez agregowanie wskazań modeli składowych. Podobnie jak w przypadku agregacyjnej metody dyskryminacji, modele składowe $\{f_j\}_{j=1}^{B}$ uzyskuje się wskutek zastosowania metody SVM do B prób uczących $D_1,\ldots,D_B$ wyodrębnionych ze zbioru uczącego D. Liczebność każdej z prób jest równa liczebności zbioru D, a elementy próby są losowane ze zbioru D według schematu losowania ze zwracaniem, ze stałym i jednakowym prawdopodobieństwem wejścia obserwacji do próby równym $\frac{1}{N}$, gdzie N – liczba obserwacji w zbiorze D. Predykcja modelu zagregowanego powstaje przez uśrednienie wskazań modeli składowych:

$$\mathrm{f}^{*}(\mathrm{x}) = \frac{1}{\mathrm{B}}\sum_{\mathrm{j}=1}^{\mathrm{B}} \mathrm{f}_{\mathrm{j}}(\mathrm{x}). \tag{1}$$

Przedstawiona i wykorzystana w niniejszym artykule metoda łączenia modeli przez rozszerzanie jest jednym z wielu podejść w zagadnieniach agregacji. Metodom agregowania (łączenia) wyników metod klasyfikacji i regresji jest poświęcona obszerna praca Gatnara [4].

1. Metoda wektorów nośnych w regresji

Zarówno w przypadku dyskryminacji, jak i regresji w algorytmie metody wektorów nośnych jest wyznaczana funkcja w postaci liniowej. Nieliniowość metody wynika z tego, że obserwacje ze zbioru uczącego są transformowane do nowej przestrzeni o dużo większym wymiarze za pomocą nieliniowego przekształcenia φ. Tak więc w przypadku danego zbioru uczącego $\left\{\left(\mathbf{x}^{\mathbf{1}}, y^{1}\right),\ldots,\left(\mathbf{x}^{\mathbf{N}}, y^{N}\right)\right\}$, gdzie $\mathbf{x}^{\mathbf{i}} \in \mathbf{R}^{d}$ oraz $y^{i} \in \mathbf{R}$ dla $i = 1,\ldots,N$, oraz nieliniowej transformacji $\varphi : \mathbf{R}^{d} \to \mathbf{Z}$ zagadnienie regresji będzie polegać na wyznaczeniu w nowej przestrzeni cech $\mathbf{Z}$ funkcji liniowej w postaci:

$$y = f(\mathbf{x}, \boldsymbol{\beta}) = \boldsymbol{\beta} \cdot \varphi(\mathbf{x}) + \beta_0, \tag{2}$$

gdzie $\varphi(\mathbf{x}^{\mathbf{i}}) \in \mathbf{Z}$ oraz $y^i \in \mathbf{R}$ dla $i = 1, \ldots, N$.

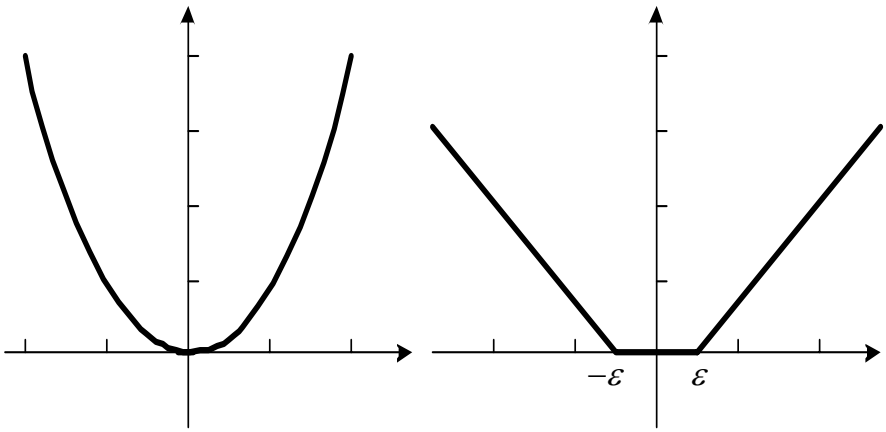
Równanie:

$$\boldsymbol{\beta} \cdot \varphi(\mathbf{x}) + \beta_0 = 0 \tag{3}$$

definiuje hiperpłaszczyznę, o której w dyskryminacji zakłada się, że ma być optymalną hiperpłaszczyzną rozdzielającą klasy, tj. położoną tak, aby rozdzielała klasy oraz by symetryczne otoczenie hiperpłaszczyzny, niezawierające żadnej obserwacji, miało jak największy promień. Tak postawiony problem można sformułować w postaci zadania optymalizacji wypukłej z liniowymi ograniczeniami [2; 8]). W regresji jest również wyznaczana optymalna hiperpłaszczyzna, lecz nie z warunkiem rozdzielania klas, tylko założeniem, że punkty przetransformowanego zbioru uczącego leżą jak najbliżej poszukiwanej hiperpłaszczyzny (por. rysunek 2). Ten geometrycznie postawiony warunek można przedstawić analitycznie jako poszukiwanie funkcji postaci (2), która minimalizuje wartość funkcjonału mierzącego jakość dopasowania funkcji do danych ze zbioru uczącego. Ów pomiar dopasowania modelu odbywa się z wykorzystaniem tzw. funkcji straty. Najczęściej w zadaniach regresji stosuje się kwadratową funkcję straty w postaci $L(y, f(\mathbf{x}, \boldsymbol{\beta})) = (y - f(\mathbf{x}, \boldsymbol{\beta}))^2$, która stanowi podstawę klasycznej metody najmniejszych kwadratów. Twórca metody Vladimir Vapnik [10] do pomiaru dopasowania zaproponował funkcję straty niewrażliwą na odchylenia rzędu $\varepsilon > 0$:

$$L^{\varepsilon}(y, f(\mathbf{x}, \boldsymbol{\beta})) = \begin{cases} 0, & \text{gdy} \quad |y - f(\mathbf{x}, \boldsymbol{\beta})| \le \varepsilon, \\ |y - f(\mathbf{x}, \boldsymbol{\beta})| - \varepsilon, & \text{gdy} \quad |y - f(\mathbf{x}, \boldsymbol{\beta})| > \varepsilon, \end{cases} \tag{4}$$

tzn. uznaje się, że dana wartość zmiennej y^i jest dobrze aproksymowana przez funkcję *f*, jeśli odległość odpowiadającego jej punktu od hiperpłaszczyzny nie jest większa niż z góry zadane $\varepsilon > 0$ (obserwacje ze zbioru uczącego znajdują się w symetrycznym otoczeniu hiperpłaszczyzny o promieniu epsylon). Różnice między kwadratową funkcją straty a funkcją Vapnika niewrażliwą na odchylenia rzędu $\varepsilon > 0$ zilustrowano na rysunku 1.



Rys. 1. Wykresy funkcji straty: po lewej funkcja kwadratowa, stosowana w metodzie najmniejszych kwadratów, po prawej funkcja Vapnika, niewrażliwa na odchylenia rzędu ε

Funkcja straty realizuje pomiar dopasowania modelu dla każdej pojedynczej obserwacji. Uwzględniając wszystkie punkty ze zbioru uczącego, otrzymano następujące kryterium jakości dopasowania i tym samym kryterium wyznaczania funkcji f:

$$\mathrm{R}_{\mathrm{emp}}(\beta) = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{L}^{\varepsilon}(\mathrm{y}^{\mathrm{i}}, \mathrm{f}(\mathrm{x}^{\mathrm{i}}, \beta)) \rightarrow \min . \tag{5}$$

Powyższe kryterium, czyli minimalizacja funkcjonału zwanego błędem empirycznym, oznacza jedynie jak najlepsze dopasowanie funkcji aproksymującej do obserwacji ze zbioru uczącego i może doprowadzić do nadmiernego dopasowania (ang. overfitting) i w konsekwencji do dużych błędów predykcji. Najlepszy model to taki, który minimalizuje nieznaną wartość błędu całkowitego:

$$R(\boldsymbol{\beta}) = \int L(y, f(\mathbf{x}, \boldsymbol{\beta})) p(\mathbf{x}, y) d\mathbf{x} dy, \tag{6}$$

jednak wartość funkcjonału (6) uwzględnia dopasowanie do wszystkich obserwacji z przestrzeni danych – również tych spoza zbioru uczącego. Na ogół badacz dysponuje informacjami cząstkowymi, zawartymi w próbie uczącej, więc wartość (6) jest niemożliwa do wyznaczenia. Vapnik i Chervonenkis udowodnili, że w przypadku liniowej funkcji regresji (definiującej hiperpłaszczyznę w przestrzeni cech), błąd całkowity (6) jest ograniczony z góry przez wartość błędu empirycznego (5) oraz długość (normę) wektora normalnego hiperpłaszczyzny [10; 2]:

Twierdzenie

Dla zbioru funkcji rzeczywistych istnieje taka stała c, że z prawdopodobieństwem $1-\delta$ *zachodzi następująca nierówność:*

$$R(\boldsymbol{\beta}) \leq \left(\frac{c}{N} \|\boldsymbol{\beta}\|^2 r^2 + R_{emp}(\boldsymbol{\beta}) \right) \log^2 N + \log \frac{1}{\delta}, \tag{7}$$

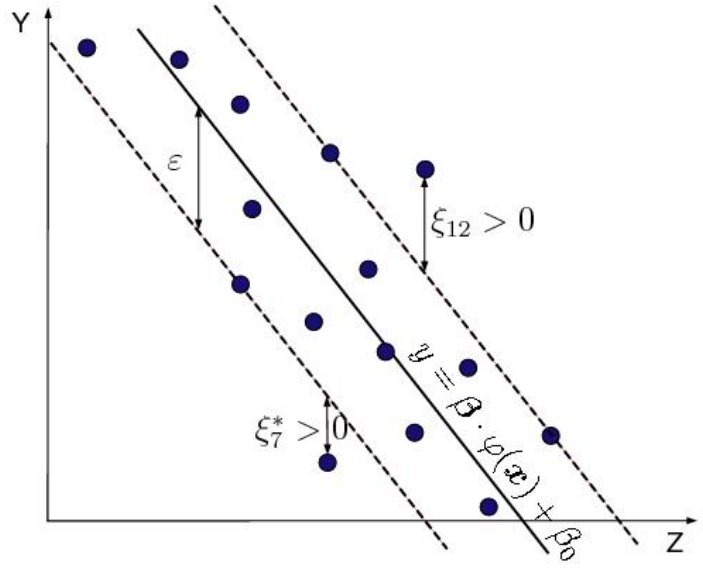
gdzie r jest promieniem hiperkuli Czebyszewa (tj. hiperkuli o najmniejszej objętości zawierającej wszystkie obserwacje ze zbioru uczącego).

Mała wartość błędu empirycznego jest związana z dobrym dopasowaniem modelu do danych ze zbioru uczącego, zaś minimalizacja normy wektora kierunkowego hiperpłaszczyzny – z lepszą zdolnością modelu do uogólniania. W metodzie wektorów nośnych jest wbudowana procedura jednoczesnej minimalizacji obu składników ograniczających błąd całkowity, która jest nazywana strukturalną minimalizacją błędu całkowitego. W algorytmie metody wektorów nośnych jest minimalizowany funkcjonał w postaci:

$$\tfrac{1}{2}\|\boldsymbol{\beta}\|^2 + C\sum_{i=1}^{N} L^{\varepsilon}(y^i, f(\mathbf{x}^i, \boldsymbol{\beta})), \tag{8}$$

gdzie parametr *C* określa kompromis między prostotą modelu a jego dopasowaniem do obserwacji ze zbioru uczącego.

W rzeczywistych zbiorach danych bardzo często występują obserwacje nietypowe, w których wartości zmiennej objaśnianej są zakłócone szumem lub różnego typu błędami. W celu uelastycznienia metody i uodpornienia jej na występowanie takiego zjawiska zostają wprowadzone zmienne $\xi_1,\ldots,\xi_N,\xi_1^*,\ldots,\xi_N^* \geq 0$. Dodatnia wartość takiej zmiennej w modelu oznacza, że odpowiadająca jej obserwacja może się znajdować poza epsylonowym otoczeniem wyznaczanej hiperpłaszczyzny. Zmniejsza to oczywiście jakość aproksymacji na zbiorze uczącym ale, szczególnie w przypadku obserwacji obarczonych błędami, może mieć pozytywny wpływ na jakość predykcji dla nowych obiektów ze zbioru rozpoznawanego.



Rys. 2. Liniowa funkcja regresji w przestrzeni **Z** i jej epsylonowe otoczenie (aproksymacja odbywa się z dokładnością do $\varepsilon > 0$) oraz konsekwencje wprowadzenia do modelu zmiennych $\xi_i, \xi_j^* \geq 0$

Geometryczne konsekwencje ustanowienia epsylonowej otoczki wokół wyznaczanej funkcji regresji w przestrzeni **Z** oraz wprowadzenia zmiennych $\xi_i, \xi_j^* \geq 0$ przedstawiono na rysunku 2.

Omówione zadanie minimalizacji z liniowymi ograniczeniami można zapisać w postaci analitycznej [7]:

$$\begin{cases} \min\limits_{\boldsymbol{\beta},\beta_0,\boldsymbol{\xi}} \frac{1}{2}\|\boldsymbol{\beta}\|^2 + C\sum\limits_{i=1}^{N}(\xi_i + \xi_i^*), \\ y^i - \left(\boldsymbol{\beta}\cdot\varphi(\mathbf{x}^i) + \beta_0\right) \leq \varepsilon + \xi_i, \quad i = 1,\ldots,N, \\ -y^i + \left(\boldsymbol{\beta}\cdot\varphi(\mathbf{x}^i) + \beta_0\right) \leq \varepsilon + \xi_i^*, \quad i = 1,\ldots,N, \\ \xi_i, \xi_i^* \geq 0. \end{cases} \tag{9}$$

Zadanie to można rozwiązać metodą mnożników Lagrange'a. Z warunku Karusha-Kuhna-Tuckera, który spełnia rozwiązanie optymalne, wynika, że wiele współczynników Lagrange'a jest równych zero. Te obserwacje, którym odpowiadają niezerowe współczynniki Lagrange'a, są nazywane wektorami nośnymi. Z geometrycznego punktu widzenia są to obserwacje leżące na brzegu epsylonowego otoczenia wyznaczonej hiperpłaszczyzny, a także te obiekty, które leżą

poza tym otoczeniem. Jedynie wektory nośne mają wpływ na położenie hiperpłaszczyzny, a tym samym na funkcję regresji [9].

O jakości modelu regresji otrzymanego metodą wektorów nośnych decyduje wybór wartości parametru C oraz przede wszystkim wybór odpowiedniej transformacji nieliniowej φ. Okazuje się, że zarówno w zadaniu optymalizacyjnym, po przejściu do jego postaci dualnej, jak i w postaci analitycznej rozwiązania przekształcenie φ występuje jedynie związane iloczynem skalarnym [5]. Wobec tej uwagi nie jest konieczne definiowanie φ wprost. Wystarczy wybrać funkcję, która będzie definiowała iloczyn skalarny **Z**, co znacznie redukuje problemy analityczne i numeryczne metody. W metodzie wektorów nośnych zamiast definiować transformację φ, użytkownik decyduje o wyborze sposobu liczenia iloczynu skalarnego w pewnej przestrzeni o większym wymiarze. Standardowo wykorzystuje się do tego celu funkcje z rodziny funkcji jądrowych w postaci $\mathrm{K}(\mathbf{u},\mathbf{v}) = \varphi(\mathbf{u}) \cdot \varphi(\mathbf{v})$, które (co udowodniono) definiują iloczyny skalarne w pewnej przestrzeni. Najczęściej stosowane funkcje jądrowe to:

1) Gaussa: $\mathrm{K}(\mathbf{u},\mathbf{v}) = \exp(-\gamma\|\mathbf{u}-\mathbf{v}\|^2)$,
2) wielomianowa: $\mathrm{K}(\mathbf{u},\mathbf{v}) = \gamma(\mathbf{u}\cdot\mathbf{v}+\delta)^d, \quad d = 1,2,\ldots,$
3) sigmoidalna: $\mathrm{K}(\mathbf{u},\mathbf{v}) = \tanh(\gamma\,\mathbf{u}\cdot\mathbf{v}+\delta)$.

Warto zaznaczyć, że wybór samej funkcji jądrowej oraz wybór wartości parametrów tej funkcji ma kluczowe znaczenie dla jakości otrzymanego modelu. Niestety brak jednoznacznej odpowiedzi na pytanie, jak dobierać rodzaj funkcji do przeprowadzanej analizy i jak identyfikować optymalne wartości parametrów. W tym celu użytkownik zazwyczaj musi przeprowadzić wstępną analizę symulacyjną dla kilku funkcji jądrowych i odpowiednio dużego zakresu wartości parametrów metody z wykorzystaniem np. sprawdzania krzyżowego. Badania empiryczne wskazują jednak na pewną przewagę wielomianowych funkcji jądrowych oraz funkcji Gaussa z odpowiednio dobranymi parametrami [8].

Przy zadanej funkcji jądrowej, definiującej iloczyn skalarny w **Z**, szukaną funkcję regresji można zapisać w postaci:

$$f(\mathbf{x}) = \sum_i (\alpha_i - \alpha_i^*)\mathrm{K}(\mathbf{x}^{\mathbf{i}}, \mathbf{x}) + \hat{\beta}_0, \tag{10}$$

gdzie α_i, α_i^* są różnymi od zera współczynnikami Langrange'a z rozwiązania optymalnego (odpowiadającymi wektorom nośnym) [2].

2. Porównanie modeli pod względem dokładności predykcji

2.1. Kryterium porównawcze

Wysoka pozycja metody wektorów nośnych w grupie metod dyskryminacji, zarówno pod względem małych błędów klasyfikacji, jak i możliwości zastosowań do różnych typów danych, wydaje się być niepodważalna [8]. Wobec faktu, że wiele własności metody SVM w dyskryminacji w sposób naturalny przenosi się na jej odpowiednik w regresji, można zadać pytanie, czy w regresji również będzie zauważalna dominująca pozycja metody wektorów nośnych. Odpowiedź na to pytanie zostanie sformułowana po przeprowadzeniu empirycznego porównania jakości różnych modeli regresji zbudowanych na zbiorach danych standardowo wykorzystywanych do badania własności i porównywania metod wielowymiarowej analizy statystycznej.

Za miarę jakości modelu przyjęto błąd średniokwadratowy. Ponieważ podstawowym celem regresji jest predykcja na nowych obiektach, spoza zbioru uczestniczącego w procesie budowy modelu, więc wartości błędów średniokwadratowych obliczono na zbiorach testowych stanowiących wyodrębnioną część 33% całego analizowanego zbioru. Pozostałe 66% tworzy zbiór uczący.

2.2. Porównywane metody

Badana grupa metod obejmuje:

1) SVM – metodę wektorów nośnych,
2) AGR_{SVM} – metodę łączenia równoległego modeli SVM przez rozszerzanie,
3) LM – klasyczną regresję liniową,
4) PPR – metodę rzutowania,
5) NNET – sieć neuronową,
6) RPART – drzewa regresyjne,
7) RFOREST – zagregowane drzewa regresyjne Breimana,
8) AGR_{RPART} – metodę łączenia równoległego drzew przez rozszerzanie.

Poszczególne metody zostały oznaczone symbolami zgodnymi z nazwami funkcji realizujących je w programie statystycznym **R**. Program ten wraz z dodatkowymi bibliotekami został wykorzystany do przeprowadzenia analizy porównawczej.

2.3. Zbiory danych

W analizie wykorzystano pięć zbiorów danych. Trzy z nich: Friedman1 (300 obserwacji, 11 zmiennych), Friedman2 (300 obserwacji, 5 zmiennych), Friedman3 (300 obserwacji, 5 zmiennych) to zbiory sztuczne, tj. takie, których

obserwacje są generowane komputerowo. Zostały one specjalnie zaprojektowane przez Friedmana [3] tak, by zawierały wiele elementów wymagających od metody odpowiedniego modelowania (nieliniowość, szum, zmienne diagnostyczne, które w ogóle nie biorą udziału w generowaniu wartości zmiennej objaśnianej itp.). Zbiory te są szeroko stosowane do porównań, podobnie jak dwa kolejne: Boston (506 obserwacji, 14 zmiennych) i Servo (167 obserwacji i 5 zmiennych). Są to zbiory danych rzeczywistych i są dostępne w internetowej bazie „UCI Repository of Machine Learning Databases" zlokalizowanej na Uniwersytecie Kalifornijskim*. Wszystkie badane zbiory i generatory zbiorów sztucznych są także dostępne w pakiecie mlbench programu statystycznego **R**.

2.4. Procedura badawcza i porównanie modeli

W pierwszym kroku z badanego zbioru w sposób losowy wybierano 33% obserwacji do zbioru testowego. Obserwacje te nie uczestniczyły w procesie budowania modelu regresji. Były jedynie wykorzystywane do wyliczenia błędu średniokwadratowego stanowiącego kryterium porównania.

Większość badanych metod wymaga od użytkownika ustalenia wartości pewnych parametrów. Symulacyjnie na zbiorze uczącym budowano wiele modeli dla różnych układów tych parametrów. Do porównania był wybierany ten model (taki układ wartości parametrów), który dawał najmniejszy błąd średniokwadratowy liczony metodą sprawdzania krzyżowego z podziałem zbioru uczącego na 10 części. Przeszukiwane zakresy parametrów dla poszczególnych metod to:

a) w metodzie wektorów nośnych wykorzystano wielomianową funkcję jądrową, zmieniając stopień wielomianu od 3 do 5 [8], wartość parametru C od 10^{-2} do 10^{2}, epsylon równe 0,1 oraz 0,5,
b) w metodzie rzutowania wartość parametru opisującego początkową liczbę funkcji składowych modelu przyjmowano na poziomie 10, 15, 20, 25, zaś końcowa liczba tych funkcji w modelu zmieniała się od 1 do 10,
c) sieć neuronowa z jedną ukrytą warstwą, z liczbą obserwacji w warstwie ukrytej zmieniającą się od 1 do ln(N),
d) dla drzew regresyjnych wymaganą minimalną liczbę obserwacji w węźle, aby nastąpił dalszy podział, ustalano na poziomie od 3 do 10, zaś kryterium minimalnej poprawy jakości modelu (przycinanie drzewa) na poziomie od 1% do 3%,
e) w metodzie zagregowanych drzew klasyfikacyjnych Breimana liczbę zmiennych losowanych przy każdym podziale ustalano na poziomie $\frac{\sqrt{d}}{2}$, $\sqrt{d}$ oraz $2\sqrt{d}$ (d – liczba zmiennych), liczbę drzew równą 100 oraz 200 [6], zaś minimalną liczbę obserwacji w liściu: 1, 5, 10.

* Dostępne przez: ftp://ftp.ics.uci.edu/pub/machine-learning-databases, http://www.ics.uci.edu/~mlearn/MLRepository.html.

2.5. Wyniki analizy

Po wyznaczeniu modelu optymalnego dla każdej z metod i każdego z badanych zbiorów, według kryterium minimalnego błędu średniokwadratowego liczonego metodą sprawdzania krzyżowego na zbiorze uczącym, obliczono błąd średniokwadratowy na zbiorze testowym i zestawiono wyniki w tabeli 1 (najlepsze wyniki zaznaczono pogrubioną czcionką).

Tabela 1

Błąd średniokwadratowy liczony na zbiorze testowym dla różnych modeli regresji

	FRIEDMAN1	FRIEDMAN2	FRIEDMAN3	BOSTON	SERVO
SVM	4,16	19206,9	0,070	15,21	21,38
AGR_{SVM}	**3,71**	20652,8	0,069	12,46	23,64
LM	8,00	38762,8	0,104	25,27	64,02
PPR	7,58	22844,2	0,026	20,04	37,96
NNET	6,69	21705,1	0,047	17,45	33,18
RPART	11,17	31007,4	0,047	18,68	21,42
AGR_{RPART}	7,64	**17170,1**	**0,025**	12,72	16,86
RFOREST	7,19	21114,9	0,027	**10,96**	**11,80**

Tylko w przypadku jednego zbioru danych metoda wektorów nośnych dała najmniejszy błąd predykcji. W pozostałych przypadkach zajmowała odpowiednio miejsca drugie, piąte i dwukrotnie trzecie. Można stąd wnioskować, że w przypadku regresji o metodzie wektorów nośnych nie można powiedzieć, że daje na ogół najlepsze rezultaty. Zdecydowanie wyniki świadczą o tym, że predykcja za pomocą modelu otrzymanego metodą SVM jest znacznie lepsza niż klasycznego modelu liniowego, lecz jednocześnie konkurencyjne wobec SVM wydają się być metody agregacyjne wykorzystujące drzewa regresyjne.

Łączenie równoległe wielu modeli zbudowanych metodą wektorów nośnych prowadzi do otrzymania modelu o mniejszym błędzie predykcji niż każdy z pojedynczych modeli składowych. Jednak nie zawsze model zagregowany jest dokładniejszy niż model SVM zbudowany od razu na całym zbiorze uczącym *D*. Okazuje się np., że dla zbiorów Friedman2 i Servo błąd zagregowanego modelu SVM jest nieco większy niż modelu SVM zbudowanego bezpośrednio na całym zbiorze uczącym *D*. Dzieje się tak, gdyż adaptacyjna metoda łączenia równoległego jest techniką redukującą przede wszystkim wariancję modelu. Metoda wektorów nośnych jest skonstruowana tak, iż jej algorytm zawiera mechanizm redukowania wariancji, dlatego może się zdarzyć, że mechanizm ten lepiej, lub przynajmniej równie dobrze, zredukuje wariancję, jak czyni to agregowanie modeli składowych. Kosztem zastosowania łączenia równoległego modeli jest cza-

sem wzrost wartości drugiej składowej błędu predykcji – obciążenia, co zostało zauważone w przypadku zbiorów Friedman2 i Servo. Ponadto zaobserwowana redukcja błędu predykcji w przypadku pozostałych zbiorów danych nie jest znacząca.

Podsumowanie

Wyniki publikowane w wielu pozycjach literatury świadczą o tym, że dyskryminacja z wykorzystaniem metody wektorów nośnych na ogół daje mniejsze błędy klasyfikacji niż metody alternatywne. Otrzymane za jej pomocą modele są nieliniowe, przestrzeń hipotez jest bardzo liczna, ale jednocześnie w metodzie jest wykorzystywany mechanizm regularyzacji przeciwdziałający nadmiernemu dopasowaniu modelu do danych ze zbioru uczącego, gdyż jest to częstą przyczyną wystąpienia dużych błędów predykcji.

Istnieje naturalny sposób przeformułowania metody wektorów nośnych tak, aby realizowała zadania regresji. Wiele pożądanych własności dyskryminacyjnej metody SVM przenosi się na jej odpowiednik regresyjny, lecz w porównaniu z innymi modelami regresji wydaje się tracić pozycję metody najdokładniejszej na rzecz metod wykorzystujących drzewa regresyjne. Ponadto metody zagregowanych drzew regresyjnych wydają się mieć przewagę w większej prostocie oraz możliwościach interpretowania modeli składowych i pozyskiwania z nich wiedzy o badanym zjawisku. Ich algorytm jest prostszy i efektywniejszy pod względem numerycznym niż w metodzie SVM.

Na uwagę zasługuje fakt, potwierdzony również przyspieszonym rozwojem badań w tym obszarze, że na ogół najlepsze rezultaty można otrzymać, gdy buduje się nie jeden model, lecz wiele modeli składowych, które się agreguje otrzymując model końcowy. Zasadę tą przede wszystkim jednak stosuje się do agregowania modeli drzew regresyjnych. Ze względu na wbudowany mechanizm redukcji wariancji modelu w algorytmie metody wektorów nośnych nie zawsze stosowanie techniki łączenia równoległego modeli prowadzi do wygenerowania modelu o dokładniejszej predykcji niż model zbudowany wprost na całym zbiorze uczącym z odpowiednio dobranymi parametrami. Wobec niewielkiej redukcji błędu predykcji oraz znacznej złożoności obliczeniowej pojedynczej metody SVM można więc uznać, że dodatkowe powiększanie tej złożoności przez stosowanie techniki łączenia równoległego wielu składowych modeli SVM nie jest zasadne. Interesująca wydaje się jednak inna możliwość łączenia wielu modeli SVM o strukturze sekwencyjnej.

Literatura

1. Breiman L.: *Bagging Predictors*.„Machine Learning" 1996, No. 24.
2. Cristianini N., Shawe-Taylor J.: *An Introduction to Support Vector Machines (and Other Kernel-based Learning Methods)*. Cambridge University Press, Cambridge 2000.
3. Friedman J.: *Multivariate Adaptive Regression Splines*. „The Annals of Statistics" 1991, 19 (1).
4. Gatnar E.: *Podejście wielomodelowe w zagadnieniach dyskryminacji i regresji*. Wydawnictwo Naukowe PWN, Warszawa 2008.
5. Gunn S.R.: *Support Vector Machines for Classification and Regression*. Technical Report, Image Speech and Intelligent Systems Research Group, University of Southampton, 1997.
6. Rozmus D.: *Random forest jako metoda agregacji modeli dyskryminacyjnych*. W: *Taksonomia 11. Klasyfikacja i analiza danych – teoria i zastosowania*. Red. K. Jajuga, M. Walesiak. Wydawnictwo Akademii Ekonomicznej, Wrocław 2004, s. 441-448.
7. Smola A., Schölkopf B.: *Learning with Kernels. Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, USA, 2002.
8. Trzęsiok M.: *Analiza wybranych własności metody dyskryminacji wykorzystującej wektory nośne*. W: *Postępy ekonometrii*. Red. A.S. Barczak. Wydawnictwo Akademii Ekonomicznej, Katowice 2004, s. 331-342.
9. Trzęsiok M.: *Metoda wektorów nośnych w konstrukcji nieparametrycznych modeli regresji*. W: *Taksonomia 12. Klasyfikacja i analiza danych*. Red. K. Jajuga, M. Walesiak. Wydawnictwo Akademii Ekonomicznej, Wrocław 2005, s. 501-510.
10. Vapnik V.: *Statistical Learning Theory*. John Wiley & Sons, Nowy Jork 1998.

BENCHMARKING AGGREGATED SUPPORT VECTOR REGRESSION MODELS

Summary

Support Vector Machines (SVM) are a state-of-the-art classification method, but they are also suitable, after a special reformulation, to perform a regression task. Similarly to classification, for a nonlinear regression problem, SVMs use the kernel trick and map the input space into a high-dimensional feature space first, and then perform linear regression in the high-dimensional feature space. One can use the model ensemble approach to try to improve the prediction accuracy. The paper presents the comparison of a single SVM, aggregated SVM and other regression models (linear regression, Projection Pursuit Regression, Neural Networks, Regression Trees, Random Forest, Bagging) by the means of a mean squared test set error.

Łukasz Wachstiel

PORÓWNANIE METOD ROZMYTEGO I PROBABILISTYCZNEGO MODELOWANIA ZJAWISKA NA PRZYKŁADZIE OCENY RYZYKA USŁUG INFORMATYCZNYCH

Wprowadzenie

Artykuł konfrontuje ze sobą dwa podejścia do oceny ryzyka w zarządzaniu usługami informatycznymi. Pierwsze z nich polega na zastosowaniu funkcji prawdopodobieństwa jako miernika gwarancji tworzonej przez daną usługę wartości. W drugim zastosowano metodę rozmytych przedziałów gwarancji, oceniając wielkość ryzyka za pomocą wartości funkcji przynależności. Celem tej konfrontacji jest zweryfikowanie następującej hipotezy: rozmyte modelowanie ryzyka jest efektywniejszą metodą od modelowania probabilistycznego przy praktycznej ocenie ryzyka usług informatycznych.

Rozważania teoretyczne zostały tu podparte analizą empiryczną z wykorzystaniem modelu oceny ryzyka opisanego przez Autora w [8], a którego krótka charakterystyka została zamieszczona w rozdziale 2.

1. Definicja ryzyka

Analiza ryzyka jest jednym z najważniejszych procesów, który należy przeprowadzać na każdym etapie cyklu życia usługi informatycznej. Począwszy od definiowania strategii, aż po proces ciągłego doskonalenia, decydenci dokonują nieustannej analizy ryzyka w mniej lub bardziej świadomy sposób. Patrząc jednak z perspektywy klienta biznesowego lub bezpośrednio od strony użytkownika krańcowego, usługa musi generować (tworzyć) wartość.

W celu wyjaśnienia pojęcia wartości usługi posłużono się znaną m.in. z metodologii ITIL® definicją, która za warunek konieczny i wystarczający wytworzenia wartości uznaje koniunkcję dwóch czynników:

– użyteczności, rozumianej jako dodatkowa korzyść dostarczana klientowi poprzez wprowadzenie usługi w celu realizacji jego potrzeb biznesowych,
– gwarancji, czyli zapewnienia funkcjonowania usługi na założonym poziomie [3].

Korzystając z powyższej definicji, można wprowadzić termin ryzyka, które będzie określane jako prawdopodobieństwo wytworzenia wartości usługi informatycznej.

Powyższe sformułowania nie dają jeszcze pełnego obrazu poszukiwanej, rzeczywistej wartości ryzyka. Można jednak dalej wnioskować, że ryzyko to inaczej zapewnienie odpowiedniej gwarancji usługi, zakładając, że każda usługa niesie ze sobą pewną określoną użyteczność. Problematyczne pozostaje pojawiające się już kilkakrotnie pojęcie gwarancji, które nie zostało dokładnie sprecyzowane.

Sięgając ponownie do biblioteki dobrych praktyk zarządzania usługami informatycznymi [3], gwarancję usługi można utożsamiać z czterema czynnikami, które wspólnie zabezpieczają jej działanie na zdefiniowanym poziomie. Są to: dostępność (availability), pojemność (capability), ciągłość (continuity), bezpieczeństwo (security).

W artykule [8] Autor podjął próbę klasyfikacji czynników ryzyka usług informatycznych ze względu na wymienione składowe gwarancji. Ze względu na duży poziom ogólności prowadzonych rozważań (nie ograniczano się do żadnej, konkretnie opisanej i sparametryzowanej usługi) wprowadzono szersze pojęcie grup ryzyka, z których każda zawierała najważniejsze czynniki mogące wpływać ujemnie na: dostępność, pojemność, ciągłość lub bezpieczeństwo usługi. Ponadto pokazano, iż każda z czterech grup posiada swojego reprezentanta, czyli element (czynnik ryzyka) mający największy wpływ na zapewnienie odpowiedniej gwarancji usługi. Wynikiem przeprowadzonych badań było stworzenie modelu oceny ryzyka usług informatycznych, którego uogólnioną postać przedstawiono w następnym punkcie.

2. Uogólniony model oceny ryzyka

Przyjęto, że okres działania usługi T można podzielić na skończoną liczbę p-parami rozłącznych podokresów w postaci $[t_i, t_{i+1}] \subset T$, gdzie $t_i < t_{i+1}$ oraz $\bigcup_{i=0}^{p} P_i = T$. Założono dodatkowo, że usługa posiada n użytkowników $U = \{u_1, u_2, \ldots, u_n\}$ oraz l zasobów $Z = \{z_1, z_2, \ldots, z_l\}$. Pomocne będzie również zdefiniowanie grup użytkowników $G = \{g_1, g_2, \ldots, g_m\}$, gdzie $m \leq n$*.

* Należy zwrócić uwagę, że słaba nierówność zakłada możliwość istnienia grup jednoelementowych.

Analiza grup czynników ryzyka doprowadziła nie tyle do wyłonienia ich reprezentantów, ile przede wszystkim do znalezienia krytycznego elementu gwarancji usługi – zasobów. Nazwano nim parametr usługi, którego zabezpieczenie ma największy wpływ na prawdopodobieństwo wytworzenia wartości końcowej. Dzięki temu można sprowadzić wielowymiarową analizę ryzyka do obserwacji stopnia wykorzystania zasobów usługi w różnych podokresach jej działania*.

Konkretyzując rozważania, można powiedzieć, że stopień wykorzystania zasobu $z \in Z$ w podokresie $P_i = [t_i, t_{i+1}] \subset T$ określa następujący zbiór rozmyty:

$$\{(t, z(t)) : t \in P_i,\ z(t) \in [0,1]\}, \tag{1}$$

$$z(t) = \min\left(\sum_{i=1}^{n} u_i(t), 1\right), \quad t \in P_i, \tag{2}$$

gdzie:

$$u_i : P_i \to [0,1],\ i = 1, \ldots, n \tag{3}$$

to procent wykorzystania zasobu *z* przez użytkownika u_i w podokresie P_i.

Następnie określa się stopień wykorzystania danego zasobu w całym okresie dostępności usługi jako średnią arytmetyczną z maksymalnych wydajności zasobów w poszczególnych podokresach z uwzględnieniem ich długości (ozn. $|P_i| = t_{i+1} - t_i$):

$$stw(z) = \frac{1}{|T|} \cdot \sum_{\substack{i=1,\ldots,p, \\ t \in P_i}} \max\left(z(t) \cdot |P_i|\right), \tag{4}$$

gdzie $|T|$ – długość całego okresu.

Na podstawie powyższych ustaleń ryzyko tworzenia wartości usługi informatycznej będzie definiowane jako gwarancja zapewnienia wydajności jej zasobów mierzonej maksymalną wartością stopni wykorzystania wszystkich zasobów obliczanych za pomocą wzoru (4).

W kolejnej części niniejszego artykułu zastosowano dwa odmienne podejścia do modelowania opisanego ryzyka w celu zweryfikowania postawionej we wprowadzeniu hipotezy.

* Stopień wykorzystania zasobu będzie również nazywany wydajnością zasobu.

3. Modelowanie ryzyka – różne podejścia

W klasycznym podejściu do oceny ryzyka zakłada się, że ma się do dyspozycji pewną funkcję prawdopodobieństwa, dzięki której wprost można policzyć wartość wystąpienia sytuacji niepożądanej. Już w pierwszej połowie XX wieku właśnie tak interpretował ryzyko Knight [2], rozróżniając je od niepewności – czynnika niemierzalnego lub niepoliczalnego. Trudno jednak takie podejście stosować, opisując zjawiska otaczającej nas rzeczywistości. Najczęściej ludzie w mowie codziennej posługują się następującymi zwrotami: „małe ryzyko", „duże ryzyko", „niewielkie ryzyko", które bardziej niosą ze sobą pewną informacje jakościową – subiektywną niż ilościową – obiektywną. Warto jednak czasami się zastanowić, czy otrzymana z pewnym przybliżeniem, a mówiąc bardziej poprawnie matematycznie – z pewnym stopniem wiarygodności informacja nie jest tak samo wartościowa, jak ta uzyskana metodami probabilistycznymi.

Idealny do konfrontacji tych dwóch alternatywnych podejść wydaje się być problem ryzyka, który już z samej definicji jest czymś bardzo subiektywnym i trudnym w rzetelnej ocenie. Stosowane do jego analizy różne miary prawdopodobieństwa dają często pozorny efekt dokładności, gdyż dane wejściowe takiego modelu bywają nierzadko standaryzowane (normalizowane), aby spełniały konkretne założenia.

Założono, że podstawową miarą ryzyka w analizowanym modelu jest opisana w rozdziale 2 gwarancja zasobów usługi, która zostanie wyznaczona na dwa sposoby.

W pierwszym z nich przyjęto, że wydajności poszczególnych zasobów są wyznaczane za pomocą funkcji gęstości rozkładu Gaussa, a następnie sprowadzane do postaci liczb rozmytych[*]. Wszystkie operacje weryfikowano jednocześnie na danych empirycznych dla uniwersyteckiej usługi poczty elektronicznej, której parametry umieszczono w tabeli 1.

Tabela 1

Ryzyko gwarancji usługi poczty elektronicznej

Podokresy dostępności	z_1	z_2	z_3	Ryzyko gwarancji
1	2	3	4	5
0-1	0,002	0,027	0,123	0,123
1-2	0,002	0,038	0,156	0,156
2-3	0,002	0,05	0,127	0,127
3-4	0,002	0,01	0,123	0,123

[*] Problem wspólnej reprezentacji danych rozważono w pracy [4]. Autorzy dowodzili, iż bardziej właściwe jest podejście, w którym zamienia się funkcje gęstości na przedziały rozmyte (a nie odwrotnie), m.in. dlatego, że nie wprowadza się pozornego złudzenia dokładności prowadzonych obliczeń.

cd. tabeli 1

1	2	3	4	5
4-5	0,002	0,002	0,167	0,167
5-6	0,003	0,004	0,189	0,189
6-7	0,002	0,057	0,182	0,182
7-8	0,017	0,079	0,155	0,155
8-9	0,031	0,231	0,183	0,231
9-10	0,033	0,32	0,137	0,32
10-11	0,042	0,367	0,189	0,367
11-12	0,069	0,412	0,199	0,412
12-13	0,105	0,519	0,199	0,519
13-14	0,145	0,51	0,199	0,51
14-15	0,201	0,421	0,199	0,421
15-16	0,189	0,412	0,198	0,412
16-17	0,078	0,49	0,125	0,49
17-18	0,023	0,321	0,134	0,321
18-19	0,011	0,284	0,167	0,284
19-20	0,007	0,164	0,102	0,164
20-21	0,008	0,079	0,105	0,105
21-22	0,004	0,082	0,162	0,162
22-23	0,003	0,045	0,198	0,198
23-0	0,004	0,033	0,134	0,134
Średnia (μ)	**0,0410417**	**0,206542**	**0,1605**	**0,206542**
Odchylenie st. (σ)	**0,0602643**	**0,185785**	**0,033052**	

Okres dostępności usługi podzielono na 24 podokresy o równej długości (godziny). Zasoby ujęto w trzech głównych grupach*: z_1 – infrastruktura aplikacji, m.in. narzędzia obsługi serwera poczty, oprogramowanie antywirusowe, narzędzia diagnostyczne; z_2 – infrastruktura sieciowa, w której analizowano przede wszystkim ruch sieciowy downstream i upstream; z_3 – infrastruktura sprzętowa, czyli głównie wydajność serwerów i ich podzespołów.

Średnie arytmetyczne (μ) oraz odchylenia standardowe (σ) uzyskane dla poszczególnych wartości wykorzystania zasobów posłużyły do wyznaczenia przedziałów ufności funkcji gęstości f rozkładu normalnego $\sim N(\mu,\sigma)$. W kolejnym etapie zamieniono przedziały ufności na α-przekroje liczb rozmytych [1], stosując następujące przekształcenia:

* Miary, jakimi się posłużono przy badaniach wydajności zasobów, można znaleźć w [6] i [7].

$$\int_{\mu-\sigma(3-k\frac{3}{\alpha})}^{\mu+\sigma(3-k\frac{3}{\alpha})} f(x)dx = \beta \quad \Leftrightarrow \quad 2\Phi(3-k\frac{3}{\alpha})-1=\beta, \tag{5}$$

$$\Phi \sim N(0,1) \tag{6}$$

gdzie Φ – dystrybuanta rozkładu normalnego.

$$z_k = \left[\mu-\sigma(3-k\frac{3}{\alpha}), \mu+\sigma(3-k\frac{3}{\alpha})\right], \tag{7}$$

$$p\left(\mu-\sigma(3-k\frac{3}{\alpha})\right) = p\left(\mu+\sigma(3-k\frac{3}{\alpha})\right) = 1-\beta, \tag{8}$$

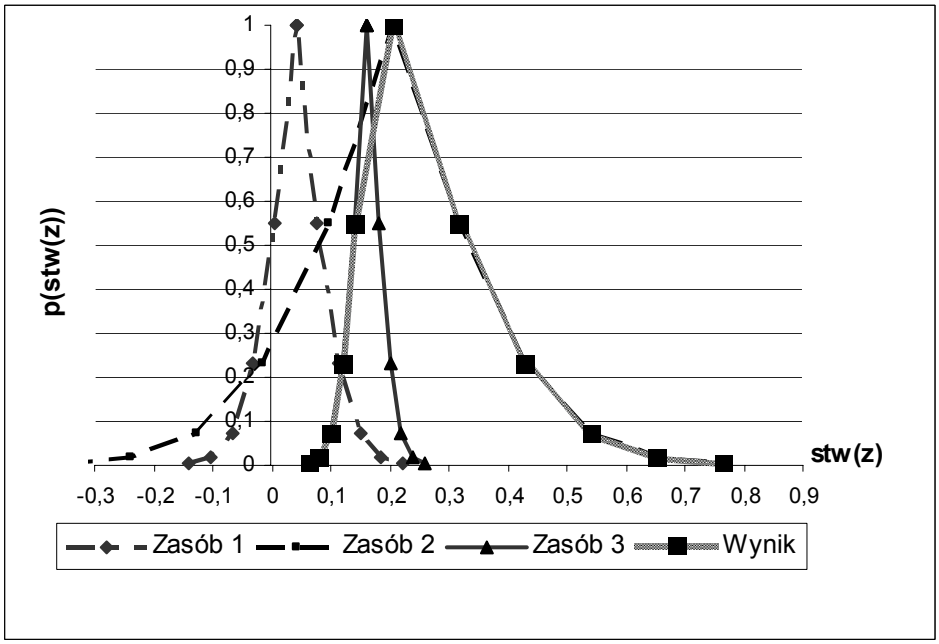
$$k \in N \cup \{0\}, \quad k \le \alpha. \tag{9}$$

Wyniki transformacji funkcji gęstości stopnia wykorzystania poszczególnych zasobów przedstawiono na rysunku 1. Dodatkowo, korzystając z operacji maksimum na przedziałach rozmytych[*], zamieszczono graficzną reprezentację funkcji przynależności dla gwarancji zasobów usługi w czasie T.

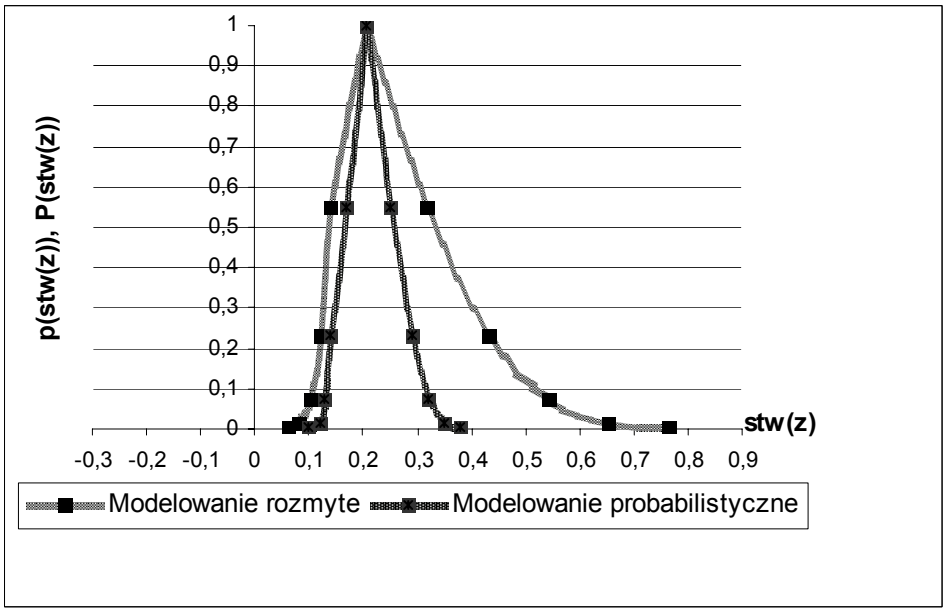
Dzięki przeprowadzonej zamianie funkcji gęstości na przedziały rozmyte przygotowano odpowiednio dane do porównania z drugą metodą modelowania ryzyka, w której stopnie wykorzystania zasobów są generowane pseudolosowo z użyciem funkcji gęstości rozkładu normalnego[**], a następnie porównywane z rzeczywistymi wartościami. Otrzymane w ten sposób wydajności poszczególnych zasobów podstawiano do wzoru (4). Wyniki modelowania rozmytego oraz symulacyjnego przedstawiono na rysunku 2.

[*] Więcej o operacjach na przedziałach liczb rozmytych można przeczytać w pracach [1; 5]. W tym miejscu użyto jednej z prostszych operacji maksimum, która wybiera maksymalne wartości liczb poszczególnych przedziałów i tworzy jeden przedział.

[**] Zastosowano do tego celu generator liczb wchodzący w skład pakietu *Mathematica*® 4.0.



Rys. 1. Transformacja funkcji gęstości do przedziałów rozmytych



Rys. 2. Porównanie modelowania rozmytego i probabilistycznego

Podsumowanie

Rezultaty przeprowadzonych symulacji (rysunek 2) wskazują na pewne podobieństwa modelowania za pomocą dwóch opisywanych metod. Widać, że poziom wykorzystania zasobów usługi o największym stopniu możliwości wystąpienia jest równy poziomowi o największym prawdopodobieństwie. Różnice pojawiają się dla wydajności o dużym stopniu niepewności. W przypadku modelowania symulacyjnego są „odrzucane" skrajne wartości wydajności, tzn. prawdopodobieństwa ich wystąpienia są bliskie zeru. Nie odzwierciedla to jednak rzeczywistej sytuacji modelowania ryzyka, w której najczęściej poszukuje się słabych punktów gwarancji usługi, chcąc je wyeliminować. Modelowanie symulacyjne rzadko uśrednia pojawiające się wartości ryzyk, wprowadzając jedynie iluzoryczną poprawność otrzymywanych wyników, która nie uwzględnia czynników mogących mieć kluczowy wpływ na funkcjonowanie usługi.

Jednocześnie można zauważyć, że niskie wartości funkcji przynależności będą zawsze implikowały niskie prawdopodobieństwa, a co za tym idzie – zbiór leżący pod wykresem funkcji prawdopodobieństwa będzie się zawierał w zbiorze rozmytym opisywanym funkcją przynależności.

Trudność zastosowania podejścia symulacyjnego przejawia się również w opisywaniu parametrów modelu za pomocą funkcji gęstości zmiennych losowych, co nie zawsze jest łatwym zadaniem. Wymaga to przeprowadzenia dużej ilości obserwacji (wykonywania iteracji na próbach o dużej liczności), dzięki którym będzie można wybrać najdokładniejszy rozkład prawdopodobieństwa. Istnieje jeszcze wiele innych, „technicznych" problemów, jak chociażby wykorzystywane generatory liczb pseudolosowych, o których więcej można przeczytać w opracowaniu [9].

Podsumowując, modelowanie z użyciem metod probabilistycznych nie dostarcza więcej informacji o poszukiwanej wartości ryzyka niż modelowanie rozmyte. Dodatkowo w przypadku modelowania probabilistycznego występuje niepożądane przy ocenie ryzyka zjawisko polegające na uśrednianiu (standaryzacji) rezultatów o skrajnie małym lub dużym prawdopodobieństwie wystąpienia. Sytuacja taka nie występuje podczas użycia drugiej metody, co wynika bezpośrednio z własności zbiorów rozmytych, które dostarczają dokładnej informacji o stopniu przynależności poszczególnych elementów. Weryfikując postawioną na początku artykułu hipotezę, stwierdzono ostatecznie, iż stosunek otrzymanych przy ocenie ryzyka obiektywnych wyników w odniesieniu do poziomu złożoności wykorzystanej aparatury badawczej jest większy w przypadku modelowania rozmytego, co w konsekwencji potwierdza wstępne przypuszczenie o przewadze tej metody w założonej kategorii efektywności.

Literatura

1. Drewniak J.: *Podstawy teorii zbiorów rozmytych: skrypt przeznaczony dla studentów IV i V roku matematyki*. Uniwersytet Śląski, Katowice 1984.
2. Knight F.: *Risk, Uncertainty and Profit*. London 1933.
3. Office of Government Commerce (OGC), *ITIL® Service Service Strategy*. Wydawnictwo TSO (The Stationary Office), Wielka Brytania 2007.
4. Róg P., Sewastianow P.: *Metoda rozmyto-przedziałowa a metoda Monte-Carlo w symulacji procesów produkcyjnych: porównanie*. XIV Górska Szkoła PTI, Szczyrk 2002.
5. Tyrala R.: *Comparison of Fuzzy Numbers Ranking Methods*. Prace Naukowe AJD, Seria Matematyka Scientific Issues of Jan Dlugosz University in Częstochowa, Mathematics, Czestochowa 2008.
6. Simmonds A.: *Wprowadzenie do transmisji danych*. WKŁ, Warszawa 2009.
7. Światowiak J.: *Microsoft Windows Server 2003/2008. Bezpieczenstwo środowiska z wykorzystaniem Forefront Security*. Helion, Gliwice 2010.
8. Wachstiel Ł.: *Identyfikacja czynników ryzyka w zarządzaniu usługami informatycznymi*. Materiały Krakowskiej Konferencji Młodych Uczonych, Grupa Naukowa Pro Futuro, Kraków 2011, s. 1091-1100.
9. Wieczorkowski R., Zieliński R.: *Komputerowe generatory liczb losowych*. WNT, Warszawa 1997.

COMPARISON OF FUZZY AND PROBABILISTIC MODELING WITH USING IT SERVICES RISK EVALUATION EXAMPLE

Summary

In the article, two approaches to IT services risk assessment has been described and compared. The first one is based on fuzzy methods in the opposite to the second one which uses probability. The following hypothesis has been proposed and proved by showing a practical example: fuzzy modeling is more effective than probabilistic in IT services risk assessment.

Maciej Wolny

ASPEKT SYTUACJI STATUS QUO WE WSPOMAGANIU WIELOKRYTERIALNEGO WYBORU BAZUJĄCEGO NA TEORII GIER

Wprowadzenie

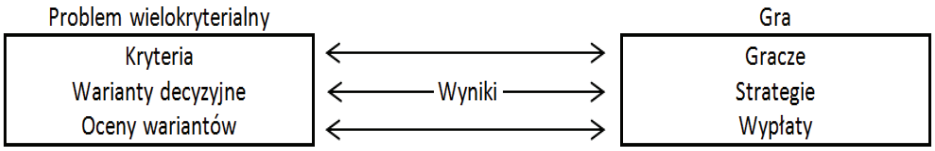
Zagadnienia wielokryterialne dotyczą sytuacji, w których rozpatruje się elementy zbioru dopuszczalnych decyzji pod kątem przynajmniej dwóch kryteriów. Wspomaganie podejmowania wielokryterialnych decyzji skupia się na następujących problematykach [1]: opisu, wyboru, sortowania i porządkowania. Problematyka wyboru polega na określeniu jednego wariantu decyzyjnego – satisfecum [2]. Przy tym racjonalne podejście implikuje, że wariant ten należy do zbioru rozwiązań sprawnych (efektywnych).

Metody wspomagania podejmowania wielokryterialnych decyzji są rozwijane od wielu lat: metody z grupy Electre [3; 4], Promethee [5], TOPSIS [6] oraz wiele innych [7; 8; 6][*].

Bez względu na poruszaną problematykę rozwiązanie wielokryterialnego problemu decyzyjnego wymaga rozpatrzenia zagadnienia porównywalności ocen (problem skali oraz normalizacji ocen) oraz przyjęcia odpowiedniej koncepcji agregacji ocen wariantów decyzyjnych [1].

Rozważania podjęte w niniejszym artykule są związane z przedstawieniem wielokryterialnego problemu decyzyjnego w postaci gry przez Maddaniego i Lunda [9]. Przy tym można zauważyć, że podejście do zagadnień wielokryterialnych z punktu widzenia teorii gier nie jest nowe. Zagadnienia wielokryterialne były formułowane jako dwuosobowa gra o sumie zerowej [10]. Analiza wielokryterialnego problemu decyzyjnego jako wieloosobowej gry niekooperacyjnej o sumie niezerowej została przedstawiona w pracy [9] oraz wcześniej w pracach [11; 12]. Punktem wyjścia do budowy modelu w postaci wieloosobowej gry jest identyfikacja związków między elementami zagadnienia wielokryterialnego a grą. Relacje te przedstawiono na rysunku 1.

[*] Cytowane metody i prace dotyczą problemów dyskretnych, najczęściej skończonych, związanych z tzw. wieloatrybutowym podejmowaniem decyzji – Multiple Attribute Decision Making. Prezentowane w artykule rozważania skupiają się wyłącznie na takich zagadnieniach.



Rys. 1. Relacje między wielokryterialnymi problemami decyzyjnymi a modelami teorii gier

Źródło: [9].

Przy budowie modelu wielokryterialnego w postaci niekooperacyjnej* gry wieloosobowej każdego gracza utożsamia się z jednym kryterium, strategie każdego z graczy są określone przez rozpatrywane warianty decyzyjne, natomiast wypłaty graczy przez oceny wariantów decyzyjnych [12]. Taka transformacja problemu implikuje konieczność ustalenia wypłat graczy w sytuacji, gdy gracze-kryteria wybierają różne strategie-warianty. Niniejszy artykuł skupia się na problemie ustalenia wyniku gry w takiej sytuacji.

Przedstawienie wielokryterialnego problemu w postaci gry nie wymaga normalizacji ocen wariantów decyzyjnych czy dodatkowych informacji dotyczących relacji między kryteriami (np. nie muszą być dane wagi poszczególnych kryteriów).

1. Sytuacja status quo

Wyborowi dopuszczalnej decyzji w sformułowanej grze odpowiada sytuacja, w której wszyscy gracze-kryteria wybierają (stosują) tę samą strategię, czyli wybierają ten sam wariant decyzyjny. Niezależność wyboru strategii przez gracza oznacza, że są możliwe sytuacje, w których przynajmniej jeden z graczy-kryteriów wybiera inny wariant.

W pracy [9] rozpatruje się zagadnienie z czterema wariantami decyzyjnymi, przy tym jeden z wariantów oznacza stan istniejący (status quo)**, natomiast są rozpatrywane dwa kryteria związane z dwiema grupami interesariuszy***. W sytuacji gdy gracze-kryteria wybiorą różne strategie-warianty, wtedy zostaje zachowany stan istniejący – wypłaty graczy w takiej sytuacji są takie same, jak

* Kooperacja między graczami-kryteriami wymaga doprowadzenia do porównywalności ocen, określenia, czy występują wypłaty uboczne – ogólnie rzecz ujmując, warunków kooperacji, które wymagają znajomości dodatkowych informacji o preferencjach. Propozycja takiego podejścia została przedstawiona w pracy [12]. Niniejszy artykuł dotyczy wyłącznie zagadnień niekooperacyjnych.

** Wybór wariantu status quo oznacza, że nie nastąpi zmiana, a pozostałe warianty są związane ze zmianą stanu istniejącego.

*** Analizowany przykład jest związany z wariantami transportu wody przez deltę rzek Sacramento oraz San Joaquin, kryteriami rozpatrywanymi są: średni roczny koszt rozwiązań oraz populacja ryb z gatunku należącego do zagrożonych – interesariuszami są więc środowisko ekologów (ochrona ekosystemu delty) oraz środowisko zainteresowane tworzeniem i zagwarantowaniem bezpiecznych dostaw wody dla Kalifornii.

w przypadku jednoczesnego wyboru wariantu status quo. Z punktu widzenia graczy-kryteriów w tak zdefiniowanej grze sytuacje te są nierozróżnialne. Model w postaci gry jest analizowany i rozwiązywany z wykorzystaniem definicji stabilności niekooperacyjnej [13]. Przedstawiona przez Madaniego i Lunda gra wymaga istnienia wariantu decyzyjnego (strategii) status quo, należy przy tym zauważyć, że jest to przypadek szczególny problemu decyzyjnego.

Z kolei w pracy [11] zaproponowano ustalenie wypłat graczy w sytuacjach, w których przynajmniej jeden z graczy wybiera inną strategię-wariant na poziomie bardzo niskim (granicznie zmierzającym do minus nieskończoności), tak aby wzmocnić przesłankę wyboru równowagi*. Rozwiązanie problemu polega na wyborze równowagi dominującej ze względu na ryzyko [14]. Zaproponowana tu koncepcja uwypukla konieczność wyboru równowagi wskazującej na wariant decyzyjny, jednak wartości wypłat graczy w sytuacji innej niż równowaga są ustalane i mogą być interpretowane wyłącznie z punktu widzenia konieczności wyboru równowagi.

Model przedstawiony przez Madaniego i Lunda stanowi z jednej strony inspirację do krytycznej analizy koncepcji przedstawionej w pracy [11], z drugiej zaś strony wyniki badań przedstawionych w pracy [11] oraz wykorzystanie koncepcji wariantu status quo są przesłanką do uogólnienia modelu przedstawionego w pracy [9].

Sytuacja status quo rozumiana w kontekście niniejszego artykułu oznacza dowolną sytuację w grze odwzorowującej wielokryterialny problem decyzyjny, w której gracze-kryteria wybierają różne strategie-warianty.

2. Budowa modelu

Niech dany będzie wielokryterialny problem decyzyjny następującej postaci:

$$\max_{x\in X} F(x) = \max_{x\in X}\left[f_1(x), f_2(x),..., f_k(x)\right], \tag{1}$$

gdzie X jest skończonym zbiorem dopuszczalnych wariantów decyzyjnych $X = \{x_1, x_2,..., x_n\}$, x – dowolnym elementem tego zbioru, f_j – j-tą funkcją-kryterium określoną na zbiorze X ($j=1,2,...,k$), $F(x)$ – wektorem grupującym wszystkie funkcje celu, $f_j(x)$ – oceną wariantu decyzyjnego względem j-tego kryterium. Ponadto są dane wszystkie oceny wariantów decyzyjnych względem wszystkich kryteriów. Rozwiązaniem problemu optymalizacji wektorowej (1) jest zbiór rozwiązań efektywnych.

* W grze występuje wiele równowag, każda równowaga odpowiada wariantowi decyzyjnemu.

Korzystając z relacji przedstawionych na rysunku 1, zagadnienie (1) można przekształcić w *k*-osobową grę niekooperacyjną o sumie niezerowej w standardowej formie:

$$G = (\Phi, H), \tag{2}$$

gdzie $\Phi = X^k$ jest zbiorem wszystkich możliwych sytuacji w grze, natomiast H jest funkcją wypłat graczy określoną na Φ. Każda sytuacja w grze jest określona jednoznacznie przez wektor strategii czystych wybranych przez każdego z graczy. Elementem zbioru Φ jest więc wektor $\phi = (x_{i1}^1, x_{i2}^2, ..., x_{ik}^k), x_{ij}^j \in X$, którego składowe oznaczają strategie poszczególnych graczy wybrane w danej sytuacji – *ij*-ta strategia jest wybierana przez *j*-tego gracza (*ij=1,2,...,n)*. Sytuację, w której wszyscy gracze wybierają strategię związaną z tym samym *i*-tym wariantem decyzyjnym, oznaczono przez:

$$\phi_i = (x_i^1, x_i^2, ..., x_i^n), x_i^1 = x_i^2 = ... = x_i^n . \tag{3}$$

Niech x^* oznacza strategię związaną z wariantem o charakterze status quo*, wtedy funkcja wypłat jest określona w następujący sposób:

$$H(\phi) = \begin{cases} (f_1(x_i), f_2(x_i), ..., f_k(x_i)) & \text{w sytuacji } \phi_i, \\ (f_1(x^*), f_2(x^*), ..., f_k(x^*)) & \text{w każdej innej sytuacji.} \end{cases} \tag{4}$$

Wyniki badań przedstawione w pracy [9] wskazują, że analiza i własności tak zdefiniowanej gry zależą od ocen wariantu status quo, który przede wszystkim musi istnieć, czyli należeć do zbioru *X*. Powstaje więc pytanie o model i rozwiązanie zagadnienia, w którym żaden z elementów zbioru rozwiązań dopuszczalnych nie ma charakteru status quo.

Wynik gry powinien jednoznacznie wskazywać na wariant decyzyjny, który jest satisfecum. Motywacją graczy-kryteriów do osiągnięcia w grze sytuacji ϕ_i, czyli jednoznacznego określenia wariantu decyzyjnego, jest punkt odniesienia, którego wypłaty odzwierciedlają sytuację, w której gracze-kryteria osiągają sytuację różną od ϕ_i – sytuację status quo. Uzyskanie koordynacji między graczami w celu osiągnięcia sytuacji ϕ_i jest możliwe, jeśli analizowana gra będzie miała charakter gry koordynacji [11]. Niech x^* zostanie redefiniowane do sytuacji status quo, która zostanie określona pesymistycznie jako potencjalny stan gry (potencjalny wariant decyzyjny), w którym gracze uzyskują minimalną możliwą wypłatę. Proponowana funkcja wypłat będzie miała wtedy następującą postać:

* Przy założeniu, że wariant status quo istnieje (w takim sensie, że brak wyboru tego samego wariantu przez wszystkich graczy-kryteriów oznacza brak zmiany stanu istniejącego).

$$H(\phi)=\begin{cases}(f_1(x_i),f_2(x_i),...,f_k(x_i)) & \text{w sytuacji } \phi_i,\\ (\min\limits_{i=1,2,..n} f_1(x_i), \min\limits_{i=1,2,..n} f_2(x_i),...,\min\limits_{i=1,2,..n} f_k(x_i)) & \text{w każdej innej sytuacji.}\end{cases} \quad (5)$$

Model zagadnienia wielokryterialnego w postaci gry (2) z funkcją wypłat (5) jest grą koordynacji, w której występuje *n* równowag w zbiorze strategii czystych. Ustalenie równowagi jest równoważne z wyborem wariantu decyzyjnego. Przyjmuje się, że tak zdefiniowana gra jest rozgrywana w „umyśle" decydenta[*], a uwzględniane przez decydenta kryteria są nieporównywalne (a przynajmniej nie jest wymagana porównywalność)[**]. Jeśli gra jest tak rozgrywana, to można ją analizować na przynajmniej dwa sposoby, uwzględniając dwa podejścia[***]:

– gra (między graczami-kryteriami) rozgrywana jednokrotnie, przy pełnej informacji o strategiach i wypłatach,
– gra rozgrywana wieloetapowo, do momentu osiągnięcia stabilnego rozwiązania (równowagi).

W przypadku rozgrywki jednorazowej do wyboru równowagi można wykorzystać koncepcję dominacji ze względu na ryzyko, natomiast w odniesieniu do gry rozgrywanej wieloetapowo – definicje stabilności niekooperacyjnych (podobnie jak to zostało przedstawione w pracach [9; 22]): w sensie Nasha[****] [15], ogólnej metaracjonalności (General Metarationality – GMR) [16], symetrycznej metaracjonalności (Symmetric Metarationality – SMR) [16], sekwencyjnej stabilności (Sequential Stability – SEQ) [17; 18], stabilności w ograniczonej liczbie ruchów (Limited Move Stability – LMS) [19; 20; 13] oraz stabilności niekrótkowzrocznej (Non-Myopic Stability – NMS) [21].

Porównanie koncepcji modelu (4) z modelem (5) zaprezentowano na powyższym przykładzie w odniesieniu do problemu rozważanego w pracy [9].

3. Przykład numeryczny – wybór sposobu transportu wody

Zagadnienie rozpatrywane w pracy Madaniego i Lunda dotyczy wyboru koncepcji transportu wody przez deltę rzek Sacramento-San Joaquin w celu zapewnienia dostaw wody dla Kalifornii przy jednoczesnej ochronie ekosystemu delty. Cele te zostały ustalone przez władze. Rozważanymi kryteriami są:

---

[*] Jeśli decydentem jest jedna osoba, jeśli problem jest wieloosobowy i decydentem jest zbiorowość, a dodatkowo z osobami są związane kryteria, to explicite można rozpatrywać zagadnienie jako grę.

[**] Jeśli kryteria są porównywalne, to powinna istnieć możliwość agregacji ocen wariantów decyzyjnych, a model problemu można zbudować bazując na teorii gier kooperacyjnych [12].

[***] Należy przy tym podkreślić, że gracze posiadają pełną informację o wszystkich wypłatach i strategiach.

[****] Równowaga Nasha związana z analizą jednokrokową, czyli jak w przypadku rozgrywania gry jednorazowo, przy czym rozwiązań stabilnych jest tyle, ile równowag w grze (nie rozpatruje się zagadnienia wyboru równowagi).

– średni roczny koszt rozwiązań (f_1),
– wielkość populacji gatunku ryb łososiowatych, który należy do gatunków zagrożonych wyginięciem (f_2).

Rozpatrywane są cztery warianty decyzyjne (rozwiązania):

– kontynuacja transportu wody przez deltę w niezmienionej formie (X_1),
– rozwiązanie tunelowe polegające na budowie np. rurociągu omijającego deltę (X_2),
– kontynuacja dotychczasowego sposobu transportu wody, jednak w mniejszym zakresie (X_3),
– zaprzestanie transportu przez deltę (X_4).

Oceny wariantów decyzyjnych są dane w postaci liczby przedziałowej (jako szacunek wielkości). W celu otrzymania oceny deterministycznej dokonuje się uśredniania ocen wariantów decyzyjnych – oblicza się średnią z końców przedziału każdej oceny wariantu decyzyjnego. Zagadnienie ogólne z ocenami wariantów (cardinal) jest przekształcane w zagadnienie uporządkowane (ordinal), wykorzystując odwrotne rangowanie – najlepszemu rozwiązaniu jest przypisana najwyższa ranga. W ten sposób przez Madaniego i Lunda został zaproponowany model w postaci deterministycznej. Model stochastyczny, który proponują autorzy, polega na iteracyjnym generowaniu oceny rozwiązania z podanego przedziału oraz rozpatrywaniu wielokrotnie modelu deterministycznego. Ostatecznie wyniki są uśredniane. Na potrzeby niniejszego artykułu będzie rozpatrywany wyłącznie model w postaci deterministycznej. Oceny wariantów decyzyjnych oraz przekształcenie ocen do macierzy gry prezentuje rysunek 2.

	f_1	f_2
X_1	2	1
X_2	4	2
X_3	3	2
X_4	1	3

status quo

$$\begin{bmatrix} (2,1) & (2,1) & (2,1) & (2,1) \\ (2,1) & (4,2) & (2,1) & (2,1) \\ (2,1) & (2,1) & (3,2) & (2,1) \\ (2,1) & (2,1) & (2,1) & (1,3) \end{bmatrix}$$

Rys. 2. Oceny wariantów decyzyjnych oraz model zagadnienia w postaci gry macierzowej z uwzględnieniem funkcji (4)

Źródło: [9].

Po lewej stronie rysunku 2 znajduje się tablica ocen wariantów decyzyjnych (transponowanych do postaci uporządkowanej)*, z prawej strony macierz wypłat

* Przekształcenie do ocen uporządkowanych nie umniejsza ogólności rozważań, ponieważ w prezentowanej analizie istotne są relacje preferencji względem każdego kryterium, a te zostały zachowane. Należy jednak zauważyć, że nastąpiła utrata informacji dotyczących proporcji odległości między wariantami – nie można stwierdzić, o ile bardziej (w jednostkach kryterium) jest preferowane jedno rozwiązanie w stosunku do drugiego.

graczy-kryteriów. Pierwsza wartość w parze to wypłata pierwszego gracza-kryterium („grającego" wierszami – wiersze odpowiadają jego strategiom postępowania), a druga wartość drugiego gracza-kryterium („grającego" kolumnami) w danej sytuacji. Macierz ta jest realizacją koncepcji funkcji wypłat (4), przy tym należy zwrócić uwagę, że nie wszystkie sytuacje (3) określające warianty decyzyjne są równowagami w sensie Nasha – tylko sytuacje odpowiadające drugiemu i trzeciemu rozpatrywanemu rozwiązaniu (druga i trzecia kolumna macierzy na przecięciu drugiego i trzeciego wiersza). Tylko te dwa rozwiązania są stabilne w sensie Nasha. W przypadku analizy jednokrotnej rozgrywki analiza powinna dotyczyć wyboru jednej z tych równowag[*].

Pozostałe definicje stabilności różnią się od siebie przede wszystkim horyzontem analizy ruchów graczy, uwzględnieniem ewentualnego pogorszenia sytuacji (wygranej) oraz posiadaną informacją o preferencjach (własnych lub wszystkich) [13; 22]. Sytuacja w grze jest stabilna (w dowolnym sensie), jeśli jest stabilna dla wszystkich graczy.

Ze względu na stabilność GMR sytuacja w grze jest analizowana w horyzoncie dwóch ruchów i jest stabilna dla *j*-tego gracza[**] wtedy i tylko wtedy, gdy każde jednostronne poprawienie sytuacji jest blokowane przez ruch innego gracza (może działać na swoją niekorzyść) do sytuacji gorszej dla gracza *j-tego*. Analizując sytuację (X_2,X_2) w rozpatrywanej grze można zauważyć, że żaden z graczy nie może polepszyć jednostronnie swojej wypłaty, dlatego sytuacja ta jest stabilna w sensie GMR. Sytuacja (X_4,X_4) jest natomiast stabilna dla gracza drugiego, ale dla gracza pierwszego nie jest stabilna, ponieważ może jednostronnie polepszyć swoją sytuację, a polepszenie nie jest blokowane przez jakikolwiek ruch oponenta (w każdej innej sytuacji gracz pierwszy otrzymuje większą wypłatę). Stabilnymi równowagami w sensie GMR w analizowanej grze są sytuacje odpowiadające drugiemu i trzeciemu wariantowi decyzyjnemu oraz sytuacja (X_1,X_1).

Stabilność w sensie SMR jest analizowana podobnie jak GMR, jednak w horyzoncie trzech ruchów. Gracz *j*-ty rozważa nie tylko własne możliwości ruchu, ale także reakcje innego gracza oraz swoje możliwości reakcji na ruch oponenta. Analizując w ten sposób grę, otrzymuje się takie same wyniki, jak w przypadku GMR.

Stabilność w sensie SEQ jest ograniczoną wersją stabilności GMR (podzbiór GMR), w której oponent może odpowiedzieć na jednostronne polepszenie

---

[*] Wariant X_2 dominuje ze względu na ryzyko rozwiązanie X_3. Ogólny schemat porównania dwóch równowag w grze dwuosobowej (do której redukuje się analizowane zagadnienie) przedstawiono w pracy [23, s. 42].

[**] Pierwszy ruch należy do *j*-tego gracza, z którego punktu widzenia rozpatruje się stabilność rozwiązania.

i-tego gracza przez wiarygodny ruch (jednostronne polepszenie, a nie dowolny ruch nawet pogarszający). Grę rozpatruje się w horyzoncie dwóch ruchów. W badanym zagadnieniu sytuacje odpowiadające wariantowi pierwszemu, drugiemu i trzeciemu cechuje stabilność w sensie SEQ, a w przypadku czwartego wariantu występuje brak stabilności.

Horyzont antycypacji przy analizie stabilności rozwiązań w sensie LMS obejmuje *h* ruchów. Stwierdza się, że sytuacja jest L_h stabilna dla *j*-tego gracza, gdy uznaje się ją za stabilną w *h* ruchach. Określenie, czy sytuacja jest stabilna, jest możliwe dzięki analizie gry w rozszerzonej formie (ekstensywnej). Zakłada się, że gracze są racjonalni i działają optymalnie – gracze mogą zrobić jednostronny ruch, jeśli są pewni powiększenia swoich wypłat (wygranych). Oznacza to, że *j*-ty gracz może celowo wykonać ruch pomniejszający wypłatę, jeśli racjonalnie zachowujący się oponent w kolejnych ruchach przejdzie do sytuacji bardziej preferowanej niż punkt wyjścia (odmiennie niż w poprzednich koncepcjach stabilności rozwiązań, w których *j*-ty gracz rozpatrywał ruch wyłącznie polepszający[*]). W analizowanym przykładzie z punktu widzenia gracza pierwszego można analizować racjonalnie L_2 stabilność LMS, ponieważ w dwóch ruchach jest możliwe przejście z jednego wariantu do drugiego, przy większej liczbie ruchów wynik się nie zmieni – jeśli sytuacja jest L_2 stabilna, to jest L_h (h > 2) stabilna. W analizowanym przypadku jedynie sytuacja odpowiadająca wariantowi drugiemu jest stabilna w sensie LMS dla $h \geq 2$, dla $h = 1$ stabilność ta jest tożsama stabilności Nasha.

Stabilność w sensie NMS jest szczególnym przypadkiem stabilności LMS, gdzie horyzont rozpatrywania ruchów jest nieograniczony. Proces ten nie może powrócić do pierwotnego stanu (sytuacji, której stabilność jest badana), czyli sekwencja analizowanych ruchów musi się zakończyć. Maksymalna możliwa liczba ruchów jest równa liczbie możliwych sytuacji w grze. W rozpatrywanym problemie jedynie sytuacja (X_2,X_2) jest stabilna w sensie NMS.

Podsumowanie analizy stabilności rozwiązań w rozpatrywanym przykładzie przedstawiono w tabeli 1.

Tabela 1

Podsumowanie analizy stabilności rozwiązań gry zdefiniowanej funkcją wypłat (4)

Rozwiązanie	**Występowanie stabilności**					
	Nash	**GMR**	**SMR**	**SEQ**	**LMS**	**NMS**
$\mathbf{X_1}$	Tak	Tak	Tak	Tak	Nie	Nie
$\mathbf{X_2}$	Tak	Tak	Tak	Tak	Tak	Tak
$\mathbf{X_3}$	Tak	Tak	Tak	Tak	Nie	Nie
$\mathbf{X_4}$	Nie	Nie	Nie	Nie	Nie	Nie

Źródło: Opracowanie własne na podstawie [9].

[*] Oznacza to, że równowaga w sensie Nasha nie musi być równowagą w sensie LMS.

Taka sama analiza problemu z wykorzystaniem zaproponowanej funkcji wypłat (5) prowadzi do podsumowania przedstawionego w tabeli 2. Na rysunku 3 zaprezentowano macierz wypłat gry w podobny sposób, jak na rysunku 2.

	f_1	f_2	
X_1	2	1	
X_2	4	2	
X_3	3	2	
X_4	1	3	
Min	1	1	*status quo*

$$\begin{bmatrix} (2,1) & (1,1) & (1,1) & (1,1) \\ (1,1) & (4,2) & (1,1) & (1,1) \\ (1,1) & (1,1) & (3,2) & (1,1) \\ (1,1) & (1,1) & (1,1) & (1,3) \end{bmatrix}$$

Rys. 3. Oceny wariantów decyzyjnych oraz model zagadnienia w postaci gry macierzowej z uwzględnieniem funkcji (5)

Tabela 2

Podsumowanie analizy stabilności rozwiązań gry zdefiniowanej funkcją wypłat (5)

Rozwiązanie	**Występowanie stabilności**					
	Nash	**GMR**	**SMR**	**SEQ**	**LMS**	**NMS**
$\mathbf{X_1}$	Tak	Tak	Tak	Tak	Nie	Nie
$\mathbf{X_2}$	Tak	Tak	Tak	Tak	Tak	Tak
$\mathbf{X_3}$	Tak	Tak	Tak	Tak	Nie	Nie
$\mathbf{X_4}$	Tak	Tak	Tak	Tak	Nie	Nie

Dodatkowo porównanie parami równowag (w sensie Nasha) wskazuje, że w przedstawionej na rysunku 3 grze równowaga (X_2,X_2) dominuje ze względu na pozostałe ryzyko.

Podsumowując przeprowadzoną analizę, można stwierdzić, że satisfecum jest wariant X_2 (rozwiązanie tunelowe). W obu wariantach modelu otrzymane rozwiązanie jest takie samo. Przy tym należy podkreślić, że wykorzystanie koncepcji gry z funkcją wypłat (4) wymaga istnienia wariantu decyzyjnego o charakterze status quo, a w porównaniu z koncepcją wykorzystującą funkcję wypłat (5) uwzględnia dodatkowe informacje wynikające z istnienia tego wariantu – sytuacja odpowiadająca wariantowi X_4 nie jest stabilna (nie jest równowagą) w żadnym sensie. Wielkość wypłat związanych z wariantem status quo określa minimum, które gwarantuje sobie racjonalny gracz-kryterium – czyli w kontekście wyboru jednego rozwiązania można wyłączyć z analizy wszystkie warianty gorsze od wariantu status quo względem dowolnego kryterium.

Podsumowanie

W artykule podjęto zagadnienie wspomagania wielokryterialnego wyboru bazujące na teorii gier w aspekcie sytuacji status quo. Relacje między problemem wielokryterialnym a grą z nim związaną przedstawiono na rysunku 1. Należy przy tym podkreślić, że analiza bazująca na teorii gier niekooperacyjnych nie wymaga doprowadzenia do porównywalności ocen wariantów decyzyjnych, jednak dotyczy problemów, w których występuje brak informacji o relacjach między kryteriami (przyjmuje się, że kryteria są nieporównywalne).

Przedstawiona propozycja funkcji wypłat (5) oraz analizy przeprowadzone w niniejszym artykule dotyczą sytuacji status quo zdefiniowanej jako każda sytuacja w grze (2), która odpowiada wyborowi różnych strategii (wariantów) przez graczy-kryteria. Dzięki przeprowadzonej analizie i porównaniu proponowanego podejścia z prezentowanym w pracy [9] (w wersji deterministycznej) można sformułować następujące wnioski:

– w zagadnieniu wyboru satisfecum, w którym występuje wariant o charakterze status quo, można wyłączyć te rozwiązania, których oceny względem dowolnego kryterium są niższe (mniej preferowane) od ocen wariantu status quo,
– uwzględnienie propozycji funkcji wypłat (5) w grze umożliwia zastosowanie analizy w problemach, w których nie występuje wariant status quo (w odróżnieniu od gry z funkcją wypłat (4)).

Przyjmując, że rozpatrywana gra jest rozgrywana w „umyśle" decydenta, analiza powinna obejmować przede wszystkim:

– stabilność w sensie Nasha wraz z wykorzystaniem do wyboru równowagi koncepcji dominacji ze względu na ryzyko,
– stabilność w sensie LMS (L_2) ze względu na możliwość strategicznego pogorszenia wypłat przy analizie oraz maksymalny horyzont antycypacji ruchów – w dwóch ruchach przechodzi się do sytuacji odpowiadającej innemu wariantowi decyzyjnemu.

Proponowane podejście może mieć szczególnie utylitarne znaczenie w sytuacji, gdy kryteria są nieporównywalne oraz na wczesnym etapie analizy problemu wielokryterialnego, gdy nie są znane explicite relacje między uwzględnianymi kryteriami.

Literatura

1. Roy B.: *Methodologie Multicritere d'Aide a la Decision* (*Wielokryterialne wspomaganie decyzji*). Editions Economica, Paris 1985.
2. Simon H.A.: *The Sciences of the Artificial*. MIT Press, Cambridge 1969.
3. Roy B.: *The Outranking Approach and Foundations of Electre Methods*. „Theory and Decision" 1991, 31, s. 49-73.

4. Roy B., Bouyssou D.: *Aide Multicritere a la Decision: Methodes et Cas*. Editions Economica, Paris 1993.
5. Brans J.-P., Vincke P.: *A Preference Ranking Organization Method (The Promethee Method for Multiple Criteria Decision-making)*. „Management Science" 1985, 31, s. 647-656.
6. Hwang C.-L., Yoon K.: *Multiple Attribute Decision Making: Methods and Applications. A State-of-the-art Survey*. Lecture Notes in Economics and Mathematical Systems 186, Springer-Verlag, Berlin-Heidelberg-New York 1981.
7. Greco S., Ehrgott M., Figueira J.: *Multiple Criteria Decision Analysis. State of the art Surveys*. Springer Science + Business Media, Inc., Boston 2005.
8. Nowak M.: *Interaktywne wielokryterialne wspomaganie decyzji w warunkach ryzyka. Metody i zastosowania*. AE, Katowice 2008.
9. Madani K., Lund J.R.: *A Monte-Carlo Game Theoretic Approach for Multi-Criteria Decision Making under Uncertainty*. „Advances in Water Resources" 2011, 34, s. 607-616.
10. Kofler E.: *O zagadnieniu optymalizacji wielocelowej*. „Przegląd Statystyczny" 1967, 1, s. 45-59.
11. Wolny M.: *Decision Making Problem with Two Incomparable Criteria – Game Theory Solution*. W: *Multiple Criteria Decision Making '07*. Red. T. Trzaskalik. Publisher of Karol Adamiecki University of Economics, Katowice 2008, s. 251-260.
12. Wolny M.: *Wspomaganie decyzji kierowniczych w przedsiębiorstwie przemysłowym. Wieloatrybutowe wspomaganie organizacji przestrzennej komórek produkcyjnych z zastosowaniem teorii gier*. Wydawnictwo Politechniki Śląskiej, Gliwice 2007.
13. Fang L., Hipel D.M., Kilgour D.M.: *Interactive Decision Making: The Graph Model for Conflict Resolution*. Wiley, New York 1993.
14. Harsanyi J.C., Selten R.: *A General Theory of Equilibrium Selection in Games*. MIT Press, Cambridge-London 1992.
15. Nash J.F.: *Non-cooperative Games*. „Annals of Mathematics" 1951, Vol. 54, No. 2, s. 286-295.
16. Howard N.: *Paradoxes of Rationality: Games, Metagames, and Political Behavior*. MIT Press, Cambridge 1971.
17. Fraser N.M., Hipel K.W.: *Solving Complex Conflicts*. „IEEE Transactions on Systems, Man and Cybernetics" 1979, SMC9(12), s. 805-816.
18. Fraser N.M., Hipel K.W.: *Conflicts Analysis: Models and Resolutions*. North-Holland, New York 1984.
19. Zagare F.C.: *Limited-move Equilibria in 2 x 2 Games*. „Theory and Decision" 16, s. 1-19.
20. Kilgour D.M., Hipel K.W., Fraser N.M.: *Solution Concept in Non-cooperative Games*. Large Scale Systems 6, s. 49-71.
21. Brams S.J., Wittman D.: *Nonmyopic Equilibria in 2 x 2 Games*. „Conflict Management and Peace Science" 1981, (6)1, s. 39-62.
22. Madani K., Hipel K.W.: *Non-Cooperative Stability Definitions for Strategic Analysis of Generic Water Resources Conflicts*. „Water Resources Management" 2011, 25, s. 1949-1977.
23. Malawski M., Wieczorek A., Sosnowska H.: *Konkurencja i kooperacja: Teoria gier w ekonomii i naukach społecznych*. Wydawnictwo Naukowe PWN, Warszawa 1997.

AN ASPECT OF STATUS QUO SITUATION IN MULTIPLE CRITERIA CHOICE SUPPORT BASING ON GAME THEORY

Summary

In this paper multiple criteria decision problem is considered as a noncooperative game. Each criterion is assigned with a player and every feasible (and effective) solution with a pure strategy. Players payoffs result from assessments of solutions in game state where every player chooses the same strategy-solution. In the analysed game status quo situation is a state in which at least one player-criterion chooses different strategy. The problem of payoff definition in such situation is considered in the paper. A proposal is to state pessimistic payoffs in status quo situations as minimal value of payoffs in situations related with considered solutions of multiple criteria problem (exactly multiple criteria choice problem). The proposal is compared with a deterministic model and solution and based on example from Madani and Lund's (2011) work.

Maciej Wolny

WSPOMAGANIE KOOPERACJI Z WYKORZYSTANIEM TEORII GIER I ANALIZY WIELOKRYTERIALNEJ

Wprowadzenie

Kooperacja między organizacjami ma istotne znaczenie w życiu gospodarczym. Podmioty gospodarcze lub ich poszczególne wydziały współpracują między innymi w celu większego stopnia integracji, wymiany doświadczeń lub rozwoju kompetencji, a przede wszystkim w celu osiągnięcia większych zysków – większych niż suma zysków uzyskanych z indywidualnych działalności, czyli chcą korzystać z efektu synergii.

Proces integrowania działań różnych wydziałów, czyli koordynacja – jedna z podstawowych składowych strategii [7], implikuje konieczność kooperacji różnych części organizacji. Znaczenie kooperacji w celu osiągnięcia efektu synergii jest widoczne w różnych sferach działalności gospodarczej – np. banki i spółki ubezpieczeniowe już od wielu lat łączą swe działania w celu sprzedaży szerokiej gamy instrumentów finansowych (kombinowana sprzedaż różnych produktów finansowych [21]). Podstawowym warunkiem współpracy jest jej opłacalność dla wszystkich kooperantów. Krytycznym czynnikiem warunkującym współdziałanie jest ustalenie wartości korzyści (wygranej, wypłaty), którą podmiot (gracz) osiągnie przez wejście w koalicję. Ustalony podział wygranej powinien być dla każdego z graczy racjonalny i akceptowalny.

W artykule przedstawiono model współpracy jako kooperacyjną grę z koalicjami. Nawiązanie współpracy w określonej koalicji powinno implikować podział zysków ze współpracy większy niż w każdej innej możliwej koalicji. Podział taki powinien więc być racjonalny, czyli znajdować się w rdzeniu gry. W związku z tym głównym celem pracy jest przedstawienie wspomagania kooperacji w sytuacji, gdy rdzeń jest pusty, oraz wspomaganie racjonalnego podziału wygranej między koalicjantów przez analizę wielokryterialną, dla której punktem wyjścia jest maksymalizacja wygranej przez każdego z graczy.

1. Podstawowe definicje i założenia

Grą kooperacyjną z koalicjami w postaci funkcji charakterystycznej nazywa się uporządkowaną parę $\Gamma = (N, v)$, gdzie $N = \{1,2,...,n\}$ jest zbiorem numerów graczy, v – funkcją charakterystyczną gry, która przyporządkowuje każdemu podzbiorowi $S \subseteq N$ łączną maksymalną nieujemną wypłatę $v(S)$, przy tym $v(\varnothing) = 0$. Biorąc pod uwagę fakt, że współdziałanie graczy implikuje efekt synergii, będą rozważane gry superaddytywne, czyli spełniające warunek:

$$v(S) + v(T) \leq v(S \cup T), \quad S, T \subseteq N, \quad S \cap T = \varnothing \tag{1}$$

oraz dodatkowo warunek istotności gry:

$$v(N) > \sum_{i=1}^{n} v(\{i\}). \tag{2}$$

Spełnienie warunków (1) i (2) implikuje wystąpienie efektu synergii.

Podział wygranej między graczy jest określony przez wektor:

$$x = (x_1,...,x_n) \in X \subseteq \Re_+^n, \tag{3}$$

którego składowe określają wygrane poszczególnych graczy. X jest zbiorem wszystkich możliwych podziałów.

Rdzeniem gry nazywa się zbiór wszystkich podziałów, które spełniają warunki:

– racjonalności zbiorowej:

$$\sum_{i=1}^{n} x_i = v(N), \tag{4}$$

– racjonalności indywidualnej:

$$x_i \geq v(\{i\}), \quad i = 1,2,...,n, \tag{5}$$

– racjonalności koalicyjnej:

$$\sum_{i \in S} x_i \geq v(S), \quad S \subseteq N. \tag{6}$$

Podział spełniający warunek racjonalności zbiorowej nazywa się preimputacją, natomiast podział spełniający warunki racjonalności indywidualnej i zbiorowej jednocześnie nazywa się imputacją. W artykule będą rozważane gry, w których istnieje przynajmniej jeden podział racjonalny w takim sensie, że

spełnia wszystkie trzy warunki racjonalności (4)-(6). W przeciwnym wypadku zawiązanie jednej wielkiej koalicji jest „nieopłacalne" przynajmniej dla jednego gracza. Niemniej należy zwrócić uwagę, że w sytuacji pustego rdzenia może istnieć podgra (ang. subgame)* (lub subgry, jeśli racjonalne jest utworzenie kilku koalicji), w której warunki racjonalności są spełnione, i prezentowane tu rozważania są adekwatne do analizy takiej podgry (podgier).

2. Wspomaganie kooperacji na podstawie analizy rdzenia gry

Warunki na niepusty rdzeń zostały sformułowane niezależnie przez Bondareva i Shapleya [3; 16] i są związane z liniowymi ograniczeniami dotyczącymi rdzenia gry. W grze istnieje niepusty rdzeń, jeśli równocześnie są spełnione następujące warunki:

$$v(N) \geq \sum_{S \subseteq N \setminus \{\varnothing\}} \lambda_S v(S), \quad \lambda_S \geq 0, \quad S \subseteq N \setminus \{\varnothing\}, \tag{7}$$

$$\sum_{S \subseteq N \setminus \{\varnothing\}} \lambda_S e_S = e_N \,, \tag{8}$$

gdzie e_S jest wektorem charakterystycznym koalicji S takim, że i-ta składowa jest równa 1, jeśli i-ty gracz należy do koalicji S, oraz 0, jeśli nie należy.

Wyrażenie (7) można przeformułować do następującego równoważnego warunku:

$$v(N) = \max_{\lambda_S \geq 0} \sum_{S \subseteq N \setminus \{\varnothing\}} \lambda_S v(S), \quad S \subseteq N \setminus \{\varnothing\}. \tag{9}$$

Nawet w sytuacji, gdy gra jest superaddytywna, przedstawione warunki mogą nie być spełnione, powstaje więc pytanie, jak wspomagać kooperację, aby wszyscy gracze współdziałali – zawiązali wielką koalicję. Szczególnego znaczenia temu zagadnieniu nadaje fakt, iż sumaryczny zysk z współdziałania wszystkich graczy jest największy**.

Odpowiedzią na postawione pytanie jest umożliwienie powiększenia wygranej wielkiej koalicji o wartość wynikającą z wyrażenia (9) przy jednoczesnym spełnieniu warunku (8):

$$\Delta v(N) = \max_{\lambda_S \geq 0} \sum_{S \subseteq N \setminus \{\varnothing\}} \lambda_S v(S) - v(N), \quad S \subseteq N \setminus \{\varnothing\}. \tag{10}$$

---

* Podgra jest tu rozumiana jako gra z uwzględnieniem wyłącznie graczy, którzy tworzą racjonalną koalicję w sensie warunków (4)-(6) w odniesieniu do zbioru tych graczy.

** Wynika to z warunku superaddytywności gry (1).

Wartość wyrażenia (10) jest nieujemna, a w sytuacji pustego rdzenia gry jest dodatnia. Wartość ta może być interpretowana jako koszt wspomagania kooperacji. Koszt ten w przypadku gry superaddytywnej oraz spełnienia warunku (8) powinien być opłacalny w tym sensie, że jest mniejszy niż przyrost wygranej. Powyższe stwierdzenie warunkuje opłacalność wspomagania kooperacji, umożliwia redefinicję wartości funkcji charakterystycznej oraz rozpatrywania nowej gry z niepustym rdzeniem $\hat{\Gamma} = (N, \hat{v})$, gdzie wartość funkcji charakterystycznej dla wielkiej koalicji wynosi:

$$\hat{v}(N) = v(N) + \Delta v(N), \tag{11}$$

natomiast dla pozostałych koalicji się nie zmienia, czyli:

$$\hat{v}(S) = v(S), \quad S \subset N. \tag{12}$$

3. Wspomaganie podziału wygranej z wykorzystaniem analizy wielokryterialnej

Jedną z kluczowych kwestii umożliwiających kooperację jest ustalenie podziału wygranej między graczy zawiązanej koalicji. Każdy gracz powinien otrzymać taką wygraną, aby był skłonny tworzyć koalicję; wygrana powinna należeć do rdzenia gry i spełniać warunki (4)-(6), czyli być racjonalna. Jeśli rdzeń jest zbiorem jednoelementowym, to podział tworzący rdzeń jest jedynym możliwym. W przypadku sytuacji nietrywialnej, czyli istnienia wielu podziałów tworzących rdzeń gry, należy wybrać (ustalić) jeden podział wygranej między graczy.

Do najważniejszych metod podziału wygranej należy wartość Shapleya [15] wynosząca dla *i*-tego gracza:

$$\varphi_i = \frac{1}{n!} \sum_{\substack{S \subseteq N \\ i \in S}} (s-1)!(n-s)![v(S) - v(S \setminus \{i\})], \tag{13}$$

gdzie *s* jest wielkością (liczebnością) koalicji *S*. Wartość Shapleya interpretuje się jako średnią wartość wygranej, jaką wnosi do wielkiej koalicji *i*-ty gracz, gdy kolejność tworzenia tej koalicji jest równoprawdopodobna, $\varphi = (\varphi_1, ..., \varphi_n)$ jest zawsze imputacją. Interpretacja wartości Shapleya powoduje, że jest ona bardzo atrakcyjnym sposobem ustalenia podziału – gracz otrzymuje tyle, ile średnio wnosi do koalicji.

Innym sposobem ustalenia podziału wygranej jest wartość Banzhafa [2], która dla *i*-tego gracza wynosi:

$$\beta_i = \frac{1}{2^{n-1}} \sum_{S \subseteq N \setminus \{i\}} \left[v(S \cup \{i\}) - v(S) \right]. \tag{14}$$

Jest najczęściej wykorzystywana jako miara siły gracza w koalicji i można ją również interpretować jako średnią wartość wygranej wnoszoną przez gracza tworzącego koalicję (w odróżnieniu od gracza „wchodzącego" w przypadku wartości Shapleya). Wartość Banzhafa $\beta = (\beta_1, ..., \beta_n)$ często nie jest imputacją, ponieważ nie spełnia warunku racjonalności zbiorowej (4). W celu wykorzystania tej wartości do ustalenia podziału wygranej należy ją poddać normalizacji:

$$\beta_i^* = \frac{\beta_i \cdot v(N)}{\sum_{i \in N} \beta_i}. \tag{15}$$

Kolejnym sposobem ustalenia podziału jest wykorzystanie koncepcji punktu Gately'ego [6]. Koncepcja ta realizuje postulat minimalizacji maksymalnej skłonności do zerwania wielkiej koalicji. Miarą skłonności do zerwania koalicji jest wyrażenie:

$$\gamma_i(x) = \frac{\sum_{j \neq i} x_j - v(N-i)}{x_i - v(i)}. \tag{16}$$

Poszukuje się więc podziału spełniającego kryterium:

$$\max_{i \in N} \gamma_i(x) \to \min. \tag{17}$$

Do najważniejszych należy koncepcja nucleolusa [14] realizująca postulat maksymalizacji największego zadowolenia z koalicji. W sytuacji niepustego rdzenia minimalizuje się wyrażenie*:

$$\max_{S \subseteq N} \left(v(S) - \sum_{i \in S} x_i \right) \to \min. \tag{18}$$

Nucleolus zawsze istnieje. Ponadto jeśli istnieje niepusty rdzeń, to należy do rdzenia, wobec tego stanowi bardzo atrakcyjną koncepcję podziału wygranej. Koncepcja nukleolusa może być rozpatrywana również w kontekście maksymalizowania największego zadowolenia z koalicji w odniesieniu do wartości wygranej koalicji [19], do wielkości koalicji [14], do wartości Shapleya [13] czy ze względu na skłonność do zerwania koalicji [10].

* Jeżeli dla danego minimalnego maksimum istnieje więcej niż jedna imputacja, to należy minimalizować kolejne maksima itd., aż do uzyskania pojedynczej imputacji.

Spośród wymienionych sposobów ustalenia jednego podziału unormowana wartość Banzhafa, wartość Shapleya oraz punkt Gately'ego mogą nie należeć do rdzenia gry. W takim przypadku istnieją przesłanki do niezawiązania się wielkiej koalicji, więc implikacją zaistnienia takiej sytuacji jest zmiana sposobu ustalenia podziału wygranej. Ze względu na cel niniejszego artykułu interesująca sytuacja jest taka, gdy gracze ustalą sposób podziału ad hoc implikujący podział nienależący do rdzenia. Innymi słowy „z góry" ustalą, że każdy z nich otrzyma tyle, ile np. średnio wnosi do koalicji (według wartości Shapleya lub Banzhafa), a podział ten nie spełni warunków (4)-(6). Należy wtedy ustalić podział będący w rdzeniu oraz jak najmniej różniący się od postulowanego.

Punktem wyjścia dalszych rozważań jest ujęcie podjętego problemu jako zagadnienia programowania wielokryterialnego. Podstawą jest stwierdzenie, że każdy z graczy pragnie uzyskać maksymalną wygraną, czyli każda ze składowych wektora podziału (3) jest maksymalizowana [1, s. 60]:

$$x \rightarrow \max. \tag{19}$$

Dopełnieniem wyrażenia (19) jest układ warunków ograniczających (4)-(6). Problem sprowadza się więc do zagadnienia liniowego programowania wielokryterialnego. Określenie pojedynczego podziału jest możliwe w sytuacji trywialnej lub przez skalaryzację zagadnienia przez przyjęcie jednej z kilku koncepcji agregacji kryteriów przedstawionych w różnych ujęciach i w sposób przeglądowy w pracach [5; 8; 9; 10; 11].

Jednym ze sposobów skalaryzacji jest zastosowanie koncepcji programowania celowego [4], która polega na minimalizacji odpowiednio zdefiniowanej funkcji strat, wynikających z odchyleń od postulowanych wartości rozważanych zmiennych. Niech $x^* = (x^*_1,...,x^*_n)$ będzie wektorem postulowanych wygranych graczy, reprezentującym wypłaty wynikające z zaakceptowanego sposobu (sprawiedliwego) podziału. Bazując na logice minimalizacji maksymalnego odchylenia od ustalonej wartości wygranej, minimalizowana funkcja strat może zostać zdefiniowana w następujący sposób:

$$\max_i | x^*_i - x_i | \rightarrow \min . \tag{20}$$

Ponieważ zwykle dla danego minimum istnieje większa liczba podziałów, więc należy przyjąć dodatkową heurystykę umożliwiającą wyodrębnienie jednego podziału – np. przez ustalenie hierarchii graczy według wartości postulowanych $x^* = (x^*_1,...,x^*_n)$ [*] lub zastosowanie procedury minimalizującej kolejne

[*] Ameljańczyk w pracy [1] proponuje do skalaryzacji zagadnienia (15) ustalenie leksykograficznej hierarchii graczy według wartości Shapleya.

maksima. Maksymalne odchylenie lub odchylenia, dla których jest osiągane minimum (20), stają się ograniczeniami. W dalszej kolejności poszukuje się minimum z maksymalnych pozostałych odchyleń itd., aż do określenia jednego podziału.

Do rozwiązania problemu ustalenia podziału wygranej można również wykorzystać metodę punktu referencyjnego przedstawioną w pracy Wierzbickiego [20]. Metoda punktu referencyjnego została pierwotnie zaproponowana jako metoda wspomagania decyzji wielokryterialnych, w której maksymalizuje się odpowiednio zdefiniowaną funkcję osiągnięcia:

$$\min_{i\in N} \sigma_i(x_i, x^*_i) + \varepsilon \sum_{i=1}^{n} \sigma_i(x_i, x^*_i) \to \max, \tag{21}$$

gdzie ε jest dodatnim parametrem umożliwiającym jednoznaczne wskazanie imputacji, $\sigma_i(x_i, x^*_i)$ – częściową funkcją osiągnięcia dla *i*-tego gracza określoną w następujący sposób:

$$\sigma_i(x_i, x^*_i) = \begin{cases} \dfrac{(x_i - v(\{i\}))}{(x^*_i - v(\{i\}))}, & x_i \in \left[v(\{i\}); x^*_i\right], \\ 1 + \alpha \dfrac{(x_i - x^*_i)}{(x^u_i - x^*_i)}, & x_i \in \left[x^*_i; v(N) - v(N \setminus \{i\})\right], \end{cases} \tag{22}$$

gdzie x^*_i oznacza wartość postulowaną (referencyjną) wypłaty *i*-tego gracza, α jest wystarczająco dużym, dodatnim parametrem optymalizacyjnym zapewniającym wklęsłość funkcji osiągnięcia, x^u_i – górnym ograniczeniem rdzenia gry:

$$x^u_i = v(N) - v(N \setminus \{i\}). \tag{23}$$

Omówione koncepcje zostaną przedstawione na przykładzie związanym ze wspomaganiem współdziałania autonomicznych oddziałów przedsiębiorstwa przy realizacji projektu.

4. Wspomaganie współdziałania oddziałów przedsiębiorstwa

Dany jest problem decyzyjny związany ze wspomaganiem podziału zysków wynikających z realizacji pewnego projektu przez przedsiębiorstwo. Na podstawie wstępnej analizy ustalono oczekiwane zyski z projektu przy podjęciu współpracy. Współpracować mogą trzy oddziały przedsiębiorstwa (trzech graczy), które posiadają dużą autonomię. Odpowiednie zyski z zawiązania współpracy przedstawia poniższa funkcja charakterystyczna (w jednostkach pieniężnych).

$$\begin{aligned} v(\varnothing) &= 0{,}00 \\ v(\{1\}) &= 22{,}00 \\ v(\{2\}) &= 24{,}00 \\ v(\{3\}) &= 30{,}00 \\ v(\{1{,}2\}) &= 74{,}00 \\ v(\{1{,}3\}) &= 75{,}00 \\ v(\{2{,}3\}) &= 76{,}00 \\ v(\{1{,}2{,}3\}) &= 112{,}00. \end{aligned} \tag{24}$$

Do dyspozycji naczelnego kierownictwa przedsiębiorstwa jest 10% zysków, pozostała część pozostaje do dyspozycji oddziałów[*]. Przedstawiona gra jest grą superaddytywną z pustym rdzeniem – wartość wyrażenia (10) wynosi 0,50, czyli tyle wynoszą koszty wspomagania kooperacji (dodatkowa premia) umożliwiające racjonalne zawiązanie wielkiej koalicji. Z punktu widzenia całej organizacji minimalny przyrost zysku wynosi: $v(\{1,2,3\}) - [v(\{1,2\}) + v(\{3\})] = 8{,}00$ jednostek pieniężnych, ponieważ maksymalna suma wypłat rozłącznych koalicji (z pominięciem wielkiej koalicji wszystkich graczy) wynosi 104,00 – jest to równoważne z utworzeniem koalicji $\{1,2\}$ oraz $\{3\}$. Część zysku będąca w gestii naczelnego kierownictwa wynosiłaby w takiej sytuacji 10,40 jednostek pieniężnych, natomiast w sytuacji utworzenia wielkiej koalicji 11,25[**]. Z tego zysku kierownictwo może przeznaczyć 0,50 jednostki pieniężnej na pokrycie kosztów kooperacji, czyli przeznaczyć do podziału dla koalicji utworzonej przez wszystkich graczy. Naczelne kierownictwo uzyskuje w takiej sytuacji 10,75 jednostek pieniężnych, czyli więcej o 0,35 jednostki pieniężnej. Natomiast z punktu widzenia oddziałów zwiększenie wygranej wielkiej koalicji powoduje, że powstaje nowa gra z wygranymi koalicji określonymi przez wyrażenia (11)-(12). Gra ta posiada niepusty rdzeń, więc istnieje przynajmniej jeden podział wygranej racjonalny w sensie warunków (4)-(6). Zasadne jest więc wspomaganie kooperacji przez naczelne kierownictwo przedsiębiorstwa.

Nowa postać funkcji charakterystycznej gry ma następującą postać:

[*] Jeśli wartości funkcji charakterystycznej miałyby oznaczać „czysty" zysk koalicjantów, to należałoby je przemnożyć przez 0,9 – otrzymana gra byłaby jednak równoważna strategicznie rozważanej, czyli obie gry powinny być rozgrywane w ten sam sposób.

[**] Jeśli zostałby umorzony narzut w odniesieniu do kosztów kooperacji, dodawanych do wygranej wielkiej koalicji, to wartość ta wynosiłaby 11,2 – dodatkowa premia do podziału między graczy w wysokości 0,05 jednostki pieniężnej.

$$
\begin{aligned}
v'(\varnothing) &= 0{,}00 \\
v'(\{1\}) &= 22{,}00 \\
v'(\{2\}) &= 24{,}00 \\
v'(\{3\}) &= 30{,}00 \\
v'(\{1{,}2\}) &= 74{,}00 \\
v'(\{1{,}3\}) &= 75{,}00 \\
v'(\{2{,}3\}) &= 76{,}00 \\
v'(\{1{,}2{,}3\}) &= 112{,}50.
\end{aligned}
\tag{25}
$$

W grze (24) istnieje niepusty rdzeń. Wartości: Shapleya, normalizowana Banzhafa, punkt Gately'ego oraz nukleolusa dla gry z funkcją charakterystyczną (24) przedstawiono w tabeli 1.

Tabela 1

Wartości: Shapleya, Banzhafa, punkt Gately'ego oraz nukleolus

	Nr gracza		
	1	2	3
Wartość Shapleya	35,33	36,83	40,33
Normalizowana wartość Banzhafa	35,50	36,88	40,12
Punkt Gately'ego	36,50	37,50	38,50
Nukleolus	36,50	37,50	38,50

Analiza otrzymanych wyników wskazuje, że punkt Gately'ego oraz nukleolus należą do rdzenia gry – nukleolus zawsze należy do rdzenia (jeśli istnieje niepusty), natomiast punkt Gately'ego może nie należeć do rdzenia dla $n > 3$ [18, s. 266]. Wartość Shapleya oraz normalizowana wartość Banzhafa nie należą do rdzenia, ponieważ dla obu wartości w przypadku koalicji $\{1,2\}$ nie jest spełniony warunek racjonalności koalicyjnej (6). W związku z tym powstaje konieczność wspomagania ustalenia podziału wygranej między graczy w sytuacji, gdy zostanie ustalone, że podział wygranej będzie według średniej wartości, jaką gracze wnoszą do koalicji bez względu czy w sensie wartości Shapleya (13), czy normalizowanej wartości Banzhafa (15). W takiej sytuacji wartości te stają się postulowanymi wygranymi graczy tworzących wielką koalicję.

Wartości wynikające z zastosowania koncepcji wykorzystującej logikę programowania celowego oraz metodę punktu referencyjnego przedstawiono w tabeli 2.

Tabela 2

Obliczone wartości wypłat graczy z zastosowaniem koncepcji programowania celowego oraz metody punktu referencyjnego

Nr gracza	1	2	3
Rozwiązanie z wykorzystaniem programowania celowego z punktem celu ustalonym przez:			
Wartość Shapleya	36,50	37,50	38,50
Normalizowaną wartość Banzhafa	36,50	37,50	38,50
Rozwiązanie z wykorzystaniem metody punktu referencyjnego z punktem ustalonym przez:			
Wartość Shapleya	36,50	37,50	38,50
Normalizowaną wartość Banzhafa	36,50	37,50	38,50

We wszystkich rozpatrywanych przypadkach otrzymano identyczny podział wygranej. Jest to konsekwencją tego, że rdzeń rozpatrywanej gry jest zbiorem jednoelementowym (dowodem na występowanie jednoelementowego rdzenia jest stwierdzenie, że rozwiązaniami optymalnymi programów liniowych z warunkami (4)-(6) rozpatrywanymi z funkcjami celu kolejno maksymalizującymi i minimalizującymi poszczególne składowe wektora (3) jest identyczny podział), czyli nie istnieje inny podział spełniający równocześnie warunki (4)-(6). Można również zaobserwować, że uzyskany wynik jest tożsamy z górnym ograniczeniem na rdzeń gry* (23). Rekomendowanym podziałem wygranej graczy w rozpatrywanej sytuacji jest więc wyznaczony jedyny podział, który spełnia wszystkie warunki racjonalności.

Podsumowanie

W artykule przedstawiono koncepcję wspomagania kooperacji, bazując na sytuacjach, w których problem można przedstawić w postaci wieloosobowej gry kooperacyjnej, w której rdzeń gry jest pusty, czyli nie istnieje żaden podział wygranej koalicji zawiązanej przez wszystkich graczy spełniający warunki racjonalności (4)-(6). Idea wspomagania polega na przyznaniu premii (zwiększeniu wygranej) za zawiązanie pożądanej koalicji. Premia ta musi być uzasadniona odpowiednio większymi od kosztów korzyściami z nawiązania współpracy. Punktem wyjścia proponowanej metody jest analiza warunków na istnienie niepustego rdzenia, w szczególności warunku (7). Rozważania prowadzą do określenia minimalnej wartości (10), o którą należy powiększyć wygraną wielkiej koalicji, aby istniał niepusty rdzeń. Należy przy tym zwrócić uwagę na fakt, iż przyznanie tej premii dokładnie w wartości (10) implikuje istnienie małego rdzenia gry (sytuacja idealna to jednoelementowy rdzeń), co z kolei ułatwia ustalenie racjonalnego podziału wygranej koalicji.

* Ogólnie rzecz ujmując, górne ograniczenie nie zawsze należy do rdzenia gry.

W artykule zaprezentowano również koncepcję wspomagania ustalenia podziału wygranej zawiązanej koalicji między graczy w sytuacji, gdy postulowany ad hoc sposób podziału implikowałby podział będący poza rdzeniem gry. Koncepcja ta bazuje na idei wielokryterialnego wspomagania decyzji. W rozważanym przypadku proponuje się minimalizację odchyleń od postulowanego podziału z wykorzystaniem metody punktu referencyjnego z funkcją osiągnięcia (21) lub z wykorzystaniem idei programowania celowego z odpowiednio zdefiniowaną funkcją strat (20).

Literatura

1. Ameljańczyk A.: *Optymalizacja wielokryterialna w problemach sterowania i zarządzania*. Ossolineum, Wrocław 1984.
2. Banzhaf J.: *Weighted Voiting Doesn't Work: A Mathematical Analysis*. „Rutgers Law Review" 1965, 19.
3. Bondareva O.N.: *Some Applications of Linear Programming Methods to the Theory of Cooperative Games* (w jęz. rosyjskim). „Problemy Kibernetiky" 1963, 10, s. 119-139.
4. Charnes A., Cooper W.W.: *Goal Programming and Multiple Objective Optimization*. „European Journal of Operational Research" 1977, 1, s. 1.
5. Galas Z., Nykowski I., Żółkiewski Z.: *Programowanie wielokryterialne*. PWE, Warszawa 1987.
6. Gately D.: *Sharing the Gains from Regional Cooperation: A Game Theoretic Application to Planning Investment in Electric Power*. „International Economic Review" 1974, 15.
7. Griffin R.W.: *Podstawy zarządzania organizacjami*. Wydawnictwo Naukowe PWN, Warszawa 2008.
8. Kaliszewski I.: *Wielokryterialne podejmowanie decyzji. Obliczenia miękkie dla złożonych problemów decyzyjnych*. WNT, Warszawa 2008.
9. Konarzewska-Gubała E.: *Programowanie przy wielorakości celów*. PWN, Warszawa 1980.
10. Littlechild S.C., Vaidya K.G.: *The Propensity to Disruptand the Disruption Nucleolus of a Characteristic Function Game*. „International Journal of Game Theory" 1976, 5, s. 151-161.
11. *Metody wielokryterialne na polskim rynku finansowym*. Red. T. Trzaskalik. PWE, Warszawa 2006.
12. Roy B.: *Wielokryterialne wspomaganie decyzji*. WNT, Warszawa 1990.
13. Seo F., Sakawa M.: *Multiple Criteria Decision Analysis in Regional Planning*. Reidel, Dordrecht, MA, 1987.

14. Schmeidler D.: *The Nucleolus of a Characteristic Function Game*. „SIAM Journal on Applied Mathematics" 1969, 17.

15. Shapley L.S.: *A Value for n-person Games*. „Annales of Mathematical Studies" 1953, 28.

16. Shapley L.S.: *On Balanced Sets and Cores*. „Naval Research Logistics Quarterly" 1967, 14, s. 453-460.

17. Shapley L.S., Schubik M.: *Quasi-cores in Monetary Economy with Non-convex Preferences*. „Econometrica" 1966, 34, s. 805-827.

18. Straffin P.D: *Teoria gier*. Scholar, Warszawa 2004.

19. Young H.P., Okada N., Hashimoto T.: *Cost Allocation in Water Resources Development – A Case Study of Sweden*. RR 80-32, IIASA, Laxenburg, Austria 1980.

20. Wierzbicki A.P.: *A Reference Point Approach to Coalition Games*. „Journal of Multicriteria Decision Analysis" 2005, 13.

21. Williams M.J.: *Synergy Works at American Express*. „Fortune" z 16.2.1987, s. 79-80.

COOPERATION SUPPORT BASED ON GAME THEORY AND MULTICRITERIAL ANALYSIS

Summary

In the paper an idea of cooperation supporting method based on game theory and multicriterial analysis. Because of taking into consideration a cooperation game with empty core an idea of supporting is based on Bondareva's and Shapley's conditions for non-emptyness of the core. There is defined a minimal value which should be added to payoff of great coalition. The minimal value should be profitable.

Also an idea of multicriterial support for establish winnings for the players in cooperative game is presented. The problem is considered in case when ad hoc arrangement is irrational in meaning that the allocation of winnings is outside the core of the game. The proposal is based on concept of goal programming and reference point method.

The proposed ideas are shown on example of cooperation problem of the three autonomous departments in enterprise.

Katarzyna Zeug-Żebro

WPŁYW REDUKCJI POZIOMU SZUMU LOSOWEGO METODĄ NAJBLIŻSZYCH SĄSIADÓW NA WARTOŚĆ NAJWIĘKSZEGO WYKŁADNIKA LAPUNOWA

Wprowadzenie

W analizie szeregów czasowych zakłada się, że w danych można wyodrębnić składnik deterministyczny oraz szum losowy, tj. zaburzenia, które utrudniają identyfikację struktury zjawiska. Źródłem szumu mogą być błędy wynikające z pomiaru lub zaokrągleń powstałe podczas badań. Może on również reprezentować czynniki egzogeniczne wpływające na dynamikę układu lub być konsekwencją statystycznego charakteru zjawisk. Większość technik analizy szeregów czasowych wiąże się z pewnymi formami filtrowania szumu w celu uwidocznienia części deterministycznej. Jedną z nich jest metoda najbliższych sąsiadów. Choć powstała ona z myślą o prognozowaniu, można ją również zastosować do filtracji danych.

Celem artykułu będzie ocena zastosowania redukcji poziomu szumu metodą najbliższych sąsiadów oraz wpływu tej filtracji na wartość największego wykładnika Lapunowa. W badaniach wykorzystano szeregi utworzone z cen zamknięcia WIG i WIG20 dwóch spółek notowanych na Giełdzie Papierów Wartościowych w Warszawie: Bytom i Żywiec oraz dwóch kursów walut: franka szwajcarskiego i dolara amerykańskiego. Dane obejmują okres od 01.01.1995 do 14.10.2011. Obliczenia przeprowadzono z użyciem programów napisanych przez autorkę w języku programowania Delphi, pakietu Microsoft Excel oraz TISEAN.

1. Rekonstrukcja przestrzeni stanów – metoda opóźnień

Metoda opóźnień jest jedną z metod rekonstrukcji przestrzeni stanów układu dynamicznego, która powstała na podstawie twierdzenia F. Takensa* o zanurzaniu [7]. Służy ona do wydobywania pewnych informacji zawartych w cha-

otycznych szeregach czasowych. W tym celu wykorzystuje się wektory opóźnień zwane również *d*-historiami. Wektory te przyjmują postać:

$$\hat{x}_t^d = \left(x_t, x_{t+\tau}, ..., x_{t+(d-1)\tau}\right), \tag{1}$$

gdzie d jest wymiarem zanurzenia, τ jest pewną liczbą naturalną nazwaną opóźnieniem czasowym, natomiast zmienna t przyjmuje $T = N - (d-1)\tau$ wartości (N jest długością szeregu czasowego).

Z badań Takensa wynika, że dla $d \geq 2m+1$, gdzie m jest wymiarem atraktora, a d jest wymiarem zanurzenia, przestrzeń stanów rozpięta przez zbiór d zmiennych będzie topologicznie równoważna z „oryginalną" przestrzenią*.

2. Redukcja poziomu szumu metodą najbliższych sąsiadów

Zadaniem metody najbliższych sąsiadów [3] jest podział szeregu czasowego x_t na część deterministyczną $\bar{x}_t$ i stochastyczną ξ_t:

$$x_t = \bar{x}_t + \xi_t, \tag{2}$$

gdzie ξ_t posiada szybko malejącą funkcję autokorelacji i jest nieskorelowana z $\bar{x}_t$.

W celu wyznaczenia $\bar{x}_l$, dla ustalonego l, należy rozważyć wektor opóźnień w przypadku, gdy czas opóźnień przyjmuje wartość jeden, $\tau = 1$:

$$\hat{x}_t^d = \left(x_t, x_{t+1}, ..., x_{t+(d-1)}\right),$$

którego jedną ze środkowych współrzędnych jest filtrowana obserwacja x_l, np. wektor w postaci:

– $\hat{x}^d_{l-\frac{d}{2}}$ dla parzystej wartości wymiaru zanurzenia,

– $\hat{x}^d_{l-\frac{d+1}{2}}$ dla nieparzystej wartości d.

Ustala się k najbliższych sąsiadów wektora $\hat{x}^d_{l-\frac{d}{2}}$:

$$\hat{x}^d_{v_1-\frac{d}{2}}, \hat{x}^d_{v_2-\frac{d}{2}}, ..., \hat{x}^d_{v_k-\frac{d}{2}} \quad . \tag{3}$$

Opierając się na wyznaczonych najbliższych sąsiadach, wartość deterministyczną $\bar{x}_l$ należy wyznaczyć ze wzoru:

* W sensie twierdzenia Takensa o zanurzeniu.

$$\bar{x}_l = \frac{1}{k}\sum_{i=1}^{k} x_{v_i} \ . \tag{4}$$

2.1. Miara *NRL*

Istotnym zagadnieniem dotyczącym redukcji szumu jest możliwość oceny jej skuteczności. Jednym z parametrów mierzących efektywność filtracji szeregu jest współczynnik poziomu redukcji szumu *NRL* [4]. Bada on zależność pomiędzy siłą szumu dodawanego do układu a strukturą geometryczną jego atraktora. Zależność ta polega na „pogrubianiu" atraktora i oddalaniu się bliskich sobie stanów w stopniu proporcjonalnym do siły szumu. Współczynnik ten wyraża się wzorem:

$$NRL(d) = \frac{1}{T}\sum_{i=1}^{T} m_i \Big/ \frac{1}{T}\sum_{i=1}^{T} M_i \ , \tag{5}$$

gdzie m_i i M_i oznaczają odległości od *i*-tego stanu (*d*-historii) do jego najbliższego i najdalszego sąsiada. Korzystając z powyższego kryterium, należy wybrać spośród otrzymanych szeregów taki, dla którego współczynnik *NRL* przyjmuje najmniejszą wartość.

3. Wykładniki Lapunowa

Wykładniki Lapunowa są miarą wrażliwości układu dynamicznego na zmianę warunków początkowych. Określają one średnie tempo oddalania lub zbliżania się dwóch początkowo bliskich sobie stanów w kolejnych iteracjach. Dodatnia wartość wykładnika mierzy rozciąganie się przestrzeni stanów, a zatem opisuje, w jaki sposób prognoza oparta na niedokładnym szacunku danych wyjściowych będzie odbiegać od realnego rozwoju układu. Jeżeli wykładnik ten przyjmuje ujemną wartość, mierzy zbieżność punktów przestrzeni stanów, czyli opisuje, ile czasu potrzebuje układ na powrót do stanu pierwotnego po wcześniejszym zaburzeniu. *m*-wymiarowy układ dynamiczny posiada *m* wykładników Lapunowa, które informują o zmianie odległości między bliskimi stanami względem odpowiedniego kierunku w przestrzeni stanów.

Definicja

Mówimy, że układ dynamiczny (X, f) jest wrażliwy na zmianę warunków początkowych, jeżeli istnieje $\varepsilon > 0$ takie, że dla każdego $x \in X$ oraz każdego otoczenia *U* punktu *x* istnieją $y \in U$ oraz $n \geq 1$ takie, że:

$$\left\| f^n(x) - f^n(y) \right\| > \varepsilon . \tag{6}$$

Przeanalizujmy układ dynamiczny w postaci:

$$x_{t+1} = f(x_t),\ t = 0,1,..., \tag{7}$$

gdzie $x_t, x_{t+1} \in X$, $X \subset R^m$ oznaczają stan systemu w chwilach t i $t+1$, f jest odwzorowaniem przestrzeni stanów w siebie z warunkiem początkowym $x_0 \in X$.

Definicja

Wykładniki Lapunowa określa wzór:

$$\lambda^i(x_0) = \lim_{n\to\infty} \frac{1}{n} \ln\left|\mu_i(n, x_0)\right|,\ i = 1,2,...,m,\ m \geq 1, \tag{8}$$

gdzie $\mu_i(n, x_0)$ są wartościami własnymi macierzy Jacobiego $Df^n(x_0)$ odwzorowania f^n:

$$Df^n(x_0) = Df(x_{n-1}) \cdot ... \cdot Df(x_1) Df(x_0), \tag{9}$$

gdzie $Df(x) = \left[\frac{\partial f_i}{\partial x_j}(x) \right]$, $i, j = 1,2,...,m$, f_i są składowymi odwzorowania f.

W szczególności dla $m = 1$ wykładnik Lapunowa przyjmuje postać:

$$\lambda(x_0) = \lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n} \ln\left\| f'(x_t) \right\|. \tag{10}$$

Najbardziej istotny z punktu identyfikacji chaosu jest największy wykładnik λ_1. Udowodniono następującą zależność:

$$\delta_n \approx \delta_0 e^{n\lambda_1}, \tag{11}$$

gdzie δ_0 jest średnią odległością pomiędzy dowolnymi dwoma położonymi blisko siebie stanami przestrzeni X, zaś δ_n jest odległością między ich obrazami w n-tej iteracji [8].

3.1. Największy wykładnik Lapunowa

W 1993 roku Rosenstein [6, s. 117-134], a rok później Kantz [2, s. 77-87] zaproponowali algorytm wyznaczania największego wykładnika Lapunowa. Algorytm ten testuje istnienie wykładniczej wrażliwości układu, przez co wskazuje celowość prowadzenia dalszych badań. Przebiega on według następujących kroków:

Krok 1. Wyznaczamy zbiory Z_i złożone z k najbliższych sąsiadów $\hat{x}_{i_j}^d$ wektorów opóźnień $\hat{x}_i^d = \left(x_i, x_{i+\tau}, ..., x_{i+(d-1)\tau}\right)$, spełniających warunek $\left|i - i_j\right| > t^*$, gdzie t^* jest ustaloną liczbą naturalną. Dodany warunek zwiększa prawdopodobieństwo, że znaleziony sąsiad nie będzie należał do trajektorii wektora $\hat{x}_i^d$.

Krok 2. Obliczamy:

$$d_n(i) = \frac{1}{k} \sum_{\hat{x}_{ij}^d \in Z_i} \left| x_{i+n} - x_{i_j+n} \right|, \quad i = 1,2,...,T;\ n = 0,1,...,n_{\max}, \tag{12}$$

gdzie $n_{\max}$ jest ustaloną liczbą naturalną określającą liczbę iteracji.

Krok 3. Wyznaczamy średnią z $d_n(i)$ po wszystkich d-historiach:

$$d_n = \frac{1}{T} \sum_{i=1}^{T} d_n(i). \tag{13}$$

Krok 4. Największy wykładnik Lapunowa jest współczynnikiem regresji:

$$\ln(d_n) = \ln(d_0) + \lambda_1 n. \tag{14}$$

Warto zauważyć, że wartość największego wykładnika Lapunowa w dużej mierze zależy od przyjętej metryki, wartości parametrów zrekonstruowanej przestrzeni stanów oraz liczby najbliższych sąsiadów k.

4. Przedmiot i przebieg badania

Interesujące rezultaty otrzymano dla szeregów finansowych pochodzących z rynków papierów wartościowych oraz walutowych. Ich długość pozwala na otrzymanie wiarygodnych rezultatów. Badaniu poddano szeregi finansowe* utworzone z cen zamknięcia WIG, WIG20, dwóch spółek notowanych na GPW w Warszawie, tj. Bytom, Żywiec oraz dziennych kursów franka szwajcarskiego

* Dane pochodzą z archiwum plików programu Omega dostępnych na stronie internetowej www.bossa.pl.

i dolara amerykańskiego. Przeanalizowano obserwacje, które były dziennymi logarytmicznymi stopami zwrotu:

$$R_t = \ln P_t - \ln P_{t-1}, \tag{15}$$

gdzie P_t jest ceną zamknięcia.

Tabela 1

Opis badanych szeregów czasowych

Szereg	Przedział czasowy	Liczba obserwacji
WIG	1995.01.02-2011.10.14	4209
WIG20	1995.01.02-2011.10.14	4209
Bytom	1995.01.05-2011.10.14	4061
Żywiec	1995.01.02-2011.10.14	4121
CHF	1995.01.01-2011.10.14	4244
USD	1995.01.01-2011.10.14	4244

Analiza wymienionych wyżej szeregów czasowych będzie przebiegała w następujących etapach:

1. Rekonstrukcja przestrzeni stanów – metoda opóźnień.
2. Redukcja szumu metodą najbliższych sąsiadów.
3. Obliczenie współczynnika poziomu redukcji szumu *NRL*.
4. Oszacowanie największego wykładnika Lapunowa.

Przeprowadzone badania empiryczne pozwoliły za pomocą metody opóźnień zrekonstruować przestrzeń stanów. Stosując metodę opartą na analizie funkcji autokorelacji – *ACF* [5], oszacowano czas opóźnień τ. Następnie za pomocą metody najbliższego pozornego sąsiada – *MNPS* [1, s. 3404-3411], obliczono wymiar zanurzenia *d*.

W kolejnym kroku badań zastosowano redukcję poziomu szumu metodą najbliższych sąsiadów*. Aby dokonać filtracji, ustalono wartość czasu opóźnienia, $\tau = 1$ oraz wartości dwóch parametrów:

- wymiar zanurzenia $d = 2,3,4,5,8,10,15,20$,
- promień otoczenia $\rho = 0{,}001;\ 0{,}01;\ 0{,}1$.

W celu oceny redukcji poziomu szumu metodą najbliższych sąsiadów wykorzystano miarę $NRL(i)$** dla $i = 2,3,...,10$. Tabela 2 zawiera najmniejszą wartość współczynnika *NRL* obliczoną dla wybranych szeregów finansowych

* Redukcję szumu przeprowadzono z wykorzystaniem darmowego programu TISEAN autorstwa H. Kantza i T. Schreibera.

** W celu obliczenia współczynnika *NRL* posłużono się programem autora napisanym w języku programowania Delphi.

oraz wartości wymiaru zanurzenia i promienia otoczenia, dla których wartość *NRL* była najniższa.

Tabela 2

Wartości miary *NRL* dla szeregów przefiltrowanych

Nazwa szeregu	Parametry filtracji		Miara *NRL*
	d	ρ	
WIG	2	0,1	0,000421
WIG20	2	0,1	0,000629
Bytom	2	0,1	0,001925
Żywiec	2	0,1	0,000744
CHF	4	0,1	0,000253
USD	3	0,1	0,000306

Można zauważyć, że prawie we wszystkich przypadkach wartości wymiaru zanurzenia i promienia otoczenia były takie same. Zmiana wystąpiła tylko w przypadku kursów walut.

Do oszacowania największego wykładnika Lapunowa posłużono się algorytmem Kantza i Rosensteina. W obliczeniach przyjęto liczbę sąsiadów $k = 1$ i wartość $t^* = 10$. Następnie zastosowano regresję liniową do przybliżania linią prostą wykresu zależności wartości $\ln d_n$ od numeru iteracji *n*. W celu porównania wyników badanie przeprowadzono dwukrotnie dla szeregów przed i po filtracji (tzn. dla szeregów otrzymanych dla parametrów *d* i ρ zamieszczonych w tabeli 2). W obydwu przypadkach przyjęto te same parametry rekonstrukcji przestrzeni stanów, tj. czas opóźnienia i wymiar zanurzenia. W tabeli 3 przedstawiono wyznaczone wartości parametrów rekonstrukcji przestrzeni stanów oraz wyniki szacowania wykładnika Lapunowa* dla analizowanych szeregów czasowych.

Tabela 3

Wyniki szacowania wykładnika Lapunowa dla finansowych szeregów czasowych

Szereg	Przyjęte parametry	Równanie regresji	Wykładnik Lapunowa
1	2	3	4
Bytom	$\tau = 5, d = 11$	$y = 0{,}0399x - 3{,}48402$ $R^2 = 0{,}06231$	–
Bytom_red	$\tau = 5, d = 11$	$y = 0{,}1849x - 4{,}992$ $R^2 = 0{,}5053$	0,1849

* Obliczenia przeprowadzono z wykorzystaniem programu własnego autora napisanego w języku programowania Delphi.

cd. tabeli 3

1	2	3	4
Żywiec	$\tau = 3, d = 10$	$y = 0{,}0043x - 4{,}08367$ $R^2 = 0{,}53954$	0,0043
Żywiec_red	$\tau = 3, d = 10$	$y = 0{,}1184x - 7{,}7597$ $R^2 = 0{,}1706$	-
WIG	$\tau = 2, d = 8$	$y = 0{,}0013x - 4{,}18708$ $R^2 = 0{,}60894$	0,0013
WIG_red	$\tau = 2, d = 8$	$y = 0{,}0415x - 10{,}562$ $R^2 = 0{,}3597$	0,0415
WIG20	$\tau = 3, d = 8$	$y = 0{,}0019x - 4{,}03212$ $R^2 = 0{,}39843$	0,0019
WIG20_red	$\tau = 3, d = 8$	$y = 0{,}0172x - 8{,}8689$ $R^2 = 0{,}6058$	0,0172
CHF	$\tau = 2, d = 10$	$y = 0{,}0181x - 5{,}46430$ $R^2 = 0{,}35259$	0,0181
CHF_red	$\tau = 2, d = 10$	$y = 0{,}0408x - 5{,}30543$ $R^2 = 0{,}33078$	0,0408
USD	$\tau = 2, d = 6$	$y = 0{,}0361x - 5{,}62090$ $R^2 = 0{,}33685$	0,0361
USD_red	$\tau = 2, d = 6$	$y = 0{,}0563x - 2{,}20383$ $R^2 = 0{,}38567$	0,0563

Można zauważyć, że prawie wszystkie objęte badaniem finansowe szeregi czasowe są wrażliwe na zmianę warunków początkowych. Jednak szeregi przefiltrowane w większym stopniu wykazały cechy chaotyczne. Wartości wykładnika Lapunowa otrzymane dla tych szeregów znacznie wzrosły. Najbardziej wrażliwa na zmianę warunków początkowych okazała się spółka Bytom. Co ciekawe, przed zastosowaniem metody redukcji poziomu szumu właśnie dla tej spółki nie można było oszacować wartości wykładnika Lapunowa. W przypadku szeregów nieprzefiltrowanych najwyższe wartości wykładnika Lapunowa uzyskano dla szeregów czasowych utworzonych z dziennych kursów walut.

Podsumowanie

Na podstawie wyników badania empirycznego należy stwierdzić, że identyfikacji chaosu w rzeczywistych szeregach czasowych warto poddawać również szeregi, w których zastosowano redukcję szumu. Wyznaczone wartości największego wykładnika Lapunowa dla szeregów przefiltrowanych znacznie przewyższały wartości tego wykładnika przed zastosowaniem filtracji. Jednak mimo zadowalających rezultatów nie należy traktować redukcji poziomu szumu bezkrytycznie, gdyż jej zastosowanie może spowodować zdeformowanie analizowanego sygnału, a co za tym idzie – błędną interpretację wyników.

Literatura

1. Abarbanel H.D., Brown R., Kennel M.B.: *Determining Embedding Dimension for Phase Space Reconstruction Using a Geometrical Construction*. „Physical Review A" 1992, Vol. 45(6), s. 3404-3411.
2. Kantz H.: *A Robust Method to Estimate the Maximal Lyapunov Exponent of a Time Series*. „Physical Letters A" 1994, Vol. 185(1), s. 77-87.
3. Kantz H., Schreiber T.: *Nonlinear Time Series Analysis*. Cambridge University Press, Cambridge 1997.
4. Orzeszko W.: *Identyfikacja i prognozowanie chaosu deterministycznego w ekonomicznych szeregach czasowych*. Polskie Towarzystwo Ekonomiczne, Warszawa 2005.
5. Ramsey J.B., Sayers C.L., Rothman P.: *The Statistical Properties of Dimension Calculations Using Small Data Sets: Some Economic Applications*. „International Economic Review" 1990, Vol. 31, No. 4.
6. Rosenstein M.T., Collins J.J., De Luca C.J.: *A Practical Method for Calculating Largest Lyapunov Exponents from Small Data Sets*. „Physica D" 1993, Vol. 65, s. 117-134.
7. Takens F.: *Detecting Strange Attractors in Turbulence*. W: *Lecture Notes in Mathematics*. Red. D.A. Rand, L.S. Young. Springer, Berlin 1981, s. 366-381.
8. Zawadzki H.: *Chaotyczne systemy dynamiczne*. Akademia Ekonomiczna, Katowice 1996.

EFFECT OF REDUCTION OF RANDOM NOISE BY METHOD OF THE NEAREST NEIGHBORS ON THE VALUE OF THE LARGEST LYAPUNOV EXPONENT

Summary

Since the presence of noise in the data can significantly affect the characteristics of dynamic systems, the aim of the article will be to evaluate effect of reduction of random noise by method of the nearest neighbors on the value of the largest Lyapunov exponent. The test will be conducted on the basis of the economic time series, which consist of closing prices of companies listed on the Warsaw Stock Exchange and the daily exchange rates.