ZESZYTY NAUKOWE
Szkoły Głównej Gospodarstwa Wiejskiego
w Warszawie

EKONOMIKA i ORGANIZACJA GOSPODARKI ŻYWNOŚCIOWEJ

NR 42 (2000)

ZESZYTY NAUKOWE
Szkoły Głównej Gospodarstwa Wiejskiego
w Warszawie

EKONOMIKA i ORGANIZACJA GOSPODARKI ŻYWNOŚCIOWEJ

NR 42 (2000)

Wydawnictwo SGGW
Warszawa 2000

Wydanie publikacji dofinansowane przez Komitet Badań Naukowych

Redaktor – Jan Kiryjow
Redaktor techniczny – Anna Karp

ISBN 83-7244-169-3

Druk: P.P. EVAN, ul. Pilicka 11, 02-629 Warszawa

Spis treści

Karol Kukuła
Analiza własności metody unitaryzacji zerowanej 5

Bolesław Borkowski, Hanna Dudek
Analiza zjawisk ekonomicznych za pomocą modeli wielorównaniowych 19

Joanna Kisielińska
Elementy teorii optymalizacji oraz metody rozwiązywania zadań optymalizacji nieliniowej 31

Andrzej Kluza
Model sieciowy wykorzystania maszyn rolniczych przy wykonywaniu prac agrotechnicznych 43

Stanisław Jabłonowski
Kontrakty terminowe w strategiach zabezpieczania się przed niekorzystnymi zmianami cen na Warszawskiej Giełdzie Towarowej 55

Arkadiusz Orłowski, Mariusz Bęben
Metody chaosu deterministycznego w badaniach ekonomicznych 67

Elżbieta Saganowska
Szacunek produktu krajowego brutto według 16 województw w 1998 roku 81

Maria Parlińska
Statystyczna kontrola jakości 93

Stanisław Stańko, Marcin Idzik
Zastosowanie różnych metod ilościowych w prognozowaniu zjawisk i procesów gospodarczych na podstawie szeregów czasowych, w których występują tendencja, wahania sezonowe i przypadkowe 107

Monika Krawiec
Konstrukcja i wycena opcji na towarowy kontrakt futures w warunkach polskiego rynku 123

Ewa Marta Syczewska
Testy kointegracji dla szeregów I(2), I(1) – Testy Hansena 135

Jadwiga Orylska, Liwiusz Siemianowski
Z badań nad modelowaniem i prognozowaniem wskaźników zanieczyszczenia Odry 149

Barbara Kowalczyk
Estymacja ilorazu dwóch średnich w kolejnym okresie badania 163

Aleksander Busłowski
Współczynniki techniczne a stabilność rozwiązania optymalnego zadania programowania liniowego 173

Aneta Becker
Metody mierzenia efektywności zarządzania w przedsiębiorstwie 189

Danuta Bogocz
Analiza zastosowań teorii zbiorów rozmytych w badaniach ekonomicznych na wybranych przykładach 215

Wacław Laskowski
Ocena ilościowego spożycia żywności przez rodziny polskie z wykorzystaniem ich klasyfikacji przy użyciu sieci neuronowych 227

Ludosław Drelichowski
Zastosowanie metod optymalizacyjnych w systemach logistyki agrobiznesu 239

Sylwester Smolik, Piotr Łukasiewicz
Opis zjawisk okresowych z monotonicznie zmieniającą się amplitudą 249

Sylwester Smolik
Modelowanie statystyczne zjawisk z trendem i sezonowością 263

Karol Kukuła
Zakład Statystyki Matematycznej AR w Krakowie

Analiza własności metody unitaryzacji zerowanej

Wstęp

Istnieje poważna liczba zjawisk, które można określić mianem złożone. Wszelkie porównania przestrzenne, a niekiedy również czasowe w zakresie zjawisk złożonych, stwarzają konieczność sporządzania ocen pewnych obiektów, a w dalszej kolejności budowy ich rankingów. Zjawiska złożone z natury rzeczy są charakteryzowane wieloma cechami, które mają różne miana i wykazują różne rzędy wielkości. Aby ocena zjawiska na podstawie cech go opisujących była możliwa, należy w sposób w miarę prosty dokonać przekształcenia ich wartości oryginalnych. Przekształcone zmienne są pozbawione mian i przybierają wartości zbliżonego rzędu wielkości. Takie sposoby transformacji wartości oryginalnych cech diagnostycznych nazywane są metodami normowania. Unormowane wartości zmiennych diagnostycznych mogą być poddane procesowi agregacji, co prowadzi do uzyskania konkretnych ocen zmiennej agregatowej. Wartości zmiennej agregatowej stanowią oceny obiektów ze względu na stan rozwojowy badanego zjawiska złożonego. Dysponowanie ocenami poszczególnych obiektów pozwala zbudować ich ranking, tj. układ, w którym obiekty są uporządkowane w kolejności od najlepszego do najgorszego ze względu na wartość zmiennej agregatowej, zwanej również zmienną syntetyczną.

W długim łańcuchu czynności prowadzących do pozyskania wielokryterialnych ocen obiektów oraz budowy ich rankingu szczególną rolę do odegrania ma proces normowania cech diagnostycznych za pomocą określonych metod. Istnieje wiele metod normowania zmiennych diagnostycznych. Metody te dają częstokroć dość zróżnicowane wyniki unormowań, nawet gdy dotyczą tych samych zmiennych. Zatem należy przyjąć, iż wybór metody ma istotny wpływ na oceny obiektów.

Wyniki badań symulacyjnych dotyczących wyników normowania przy zastosowaniu dziesięciu formuł (zob. [5], ss. 111–151) upoważniają do zwrócenia szczególnej uwagi na metodę unitaryzacji zerowanej, w skrócie MUZ.

Kwintesencją badań zjawisk złożonych jest ich ujęcie porównawcze, co oznacza, iż poziom zjawiska rozpatruje się w obiektach. Obiekty w początko-

wej fazie charakteryzowane są za pomocą wielu cech. Przyjmijmy, że O oznaczać będzie zbiór obiektów:

$$O = \{O_1, O_2, \ldots, O_r\}, \tag{1}$$

gdzie r jest liczbą badanych obiektów. Każdy z obiektów jest opisany przez zbiór zmiennych diagnostycznych X.

W podejściu badawczym określonym przymiotnikiem „statyczne" rozpatrywane zjawisko analizowane jest w jednym, wybranym okresie. Zwykle jest to ostatni z okresów, gwarantujący pozyskanie odpowiednich informacji statystycznych. W podejściu tym zrezygnowano z oznaczania zmiennych indeksem t. Niezbędne w badaniach wielowymiarowych dane tworzą macierz o postaci:

$$\mathbf{X} = [x_{ij}] = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1s} \\ x_{21} & x_{22} & \cdots & x_{2s} \\ \vdots & \vdots & \vdots & \vdots \\ x_{r1} & x_{r2} & \cdots & x_{rs} \end{bmatrix}, \qquad \begin{pmatrix} i = 1, \ldots, r \\ j = 1, \ldots, s \end{pmatrix}, \tag{2}$$

gdzie x_{ij} oznacza realizację zmiennej X_j w obiekcie O_i. Zatem *i*-ty obiekt charakteryzuje następujący wektor zmiennych:

$$\mathbf{X}_i = [x_{i1} \quad x_{i2} \quad \ldots \quad x_{is}], \qquad (i = 1, \ldots, r). \tag{3}$$

Wektor $\mathbf{X}_i$ jest s-wymiarową obserwacją charakteryzującą obiekt O_i. Łatwo zauważyć, iż każdemu obiektowi odpowiada punkt w przestrzeni s--wymiarowej.

Podział metod normujących

Rozpatrzenie zalet, jak również ewentualnych wad metody unitaryzacji zerowanej (MUZ) rodzi konieczność jej prezentacji na tle innych metod normujących cechy diagnostyczne w możliwie szerokim kontekście. Dla realizacji tego celu niezbędne wydają się omówienie i wszechstronna analiza własności procedur normowania najczęściej stosowanych i mających już stałe miejsce

w literaturze przedmiotu. Ze względu na znaczną ich liczbę, prezentację procedur normowania ograniczono do wybranych metod. Przedstawiono i omówiono te, które znajdują akceptację wśród badaczy wykorzystujących aparat wielowymiarowej analizy porównawczej w pracach empirycznych. Z drugiej zaś strony przy wyborze uwzględniono tradycyjny już podział procedur normowania (zob. T. Borys – [2] lub T. Grabiński – [4], s. 33–34), który obejmuje:

- metody standaryzacji,
- metody unitaryzacji,
- przekształcenia ilorazowe,
- metody rangowe.

Normowanie jest działaniem mającym na celu przysposobić zmienne diagnostyczne do roli kryteriów cząstkowych w procesie oceny zjawiska złożonego. Dodajmy, iż zwykle cechy diagnostyczne wyrażone są w różnych jednostkach miary oraz odpowiadają im zróżnicowane zakresy liczbowe. Uwzględniając potrzebę pozbycia się mian oraz ujednolicenia zakresów liczbowych zmiennych diagnostycznych, metody normujące służą transformacji bezwzględnych wartości na wartości względne. W tym sensie każda z metod normowania (oprócz metody rangowej) jest pewnym przekształceniem ilorazowym, które w końcowym wyniku daje zmienną diagnostyczną transformowaną. Zmienna ta jest pozbawiona miana i ujednolicona co do zakresu wartości, jakie może przyjmować. Powstaje zatem pewna wątpliwość, czy słuszne jest określenie „metody oparte na przekształceniu ilorazowym", do których zalicza się m.in. metodę zaproponowaną przez D. Strahl [7], E. Nowaka [6], S. Bartosiewicz [1] oraz M. Cieślak [3], skoro wszystkie wymienione metody (z metodami standaryzacyjnymi i unitaryzacyjnymi włącznie) są skonstruowane w formie ilorazowej. Oznacza to dzielenie oryginalnej wartości cechy bądź różnicy między tą wartością a określonym parametrem (średnią, minimum wszystkich wartości itp.) przez odpowiednią wartość stałą wyrażoną tą samą jednostką co zmienna oryginalna. Dlatego też, biorąc pod uwagę zaistniały już podział metod normujących (zob. [2] lub [4]), proponuję przyjąć następującą ich specyfikację:

A. Metody oparte na formule przekształcenia ilorazowego.
B. Metody rangowe.

Metody z grupy A oparte na formule przekształcenia ilorazowego przyjmują różne punkty odniesienia, które można określić jako:

1) miary zróżnicowania cech, takie jak:
   a) odchylenie standardowe zmiennej (ten punkt odniesienia wykorzystują metody standaryzacyjne),

b) rozstęp zmiennej (ten punkt odniesienia wykorzystują metody unitaryzacyjne);

2) inne parametry stałe cechy, takie jak:
   a) średnia arytmetyczna zmiennej,
   b) maksymalna wartość zmiennej,
   c) minimalna wartość zmiennej,
   d) długość wektora realizacji zmiennej,
   e) suma realizacji zmiennej.

Przy omawianiu metod normowania cech diagnostycznych skoncentrowano uwagę na metodach z grupy A. Metody rangowe zastosowane do zmiennych mierzonych na skalach ilorazowej i przedziałowej wprawdzie są możliwe do zastosowania, co tłumaczy łatwość przejścia z wyższych skal do niższych, lecz takie postępowanie należy zawsze łączyć z poważną stratą informacji. A oto wybrane formuły normujące z grupy metod A:

(I) $$z_{ij} = \frac{x_{ij} - \overline{X}_j}{S(X_j)}, \qquad S(X_j) > 0,$$

(II) $$z_{ij} = \frac{x_{ij}}{S(X_j)}, \qquad S(X_j) > 0,$$

(III) $$z_{ij} = \frac{x_{ij}}{\max\limits_i x_{ij} - \min\limits_i x_{ij}}, \qquad \max_i x_{ij} > \min_i x_{ij},$$

(IV) $$z_{ij} = \frac{x_{ij} - \overline{X}_j}{\max\limits_i x_{ij} - \min\limits_i x_{ij}}, \qquad \max_i x_{ij} > \min_i x_{ij},$$

(V) $$z_{ij} = \frac{x_{ij} - \min\limits_i x_{ij}}{\max\limits_i x_{ij} - \min\limits_i x_{ij}}, \qquad \max_i x_{ij} > \min_i x_{ij},$$

(VI) $$z_{ij} = \frac{x_{ij}}{\max\limits_i x_{ij}}, \qquad \max_i x_{ij} \neq 0,$$

(VII) $$z_{ij} = \frac{x_{ij}}{\min\limits_{i} x_{ij}}, \qquad \min\limits_{i} x_{ij} \neq 0,$$

(VIII) $$z_{ij} = \frac{x_{ij}}{\overline{X}_j}, \qquad \overline{X}_j \neq 0,$$

(IX) $$z_{ij} = \frac{x_{ij}}{\sum\limits_{i=1}^{r} x_{ij}}, \qquad \sum_{i=1}^{r} x_{ij} \neq 0,$$

(X) $$z_{ij} = \frac{x_{ij}}{\left[\sum\limits_{i=1}^{r} x_{ij}^2\right]^{0,5}}, \qquad \sum_{i=1}^{r} x_{ij}^2 > 0.$$

Szczególne miejsce poświęcono przypadkowi normowania zmiennych diagnostycznych przyjmujących wartości ze zbioru liczb rzeczywistych R (chodzi o takie zmienne, które mogą być zarówno liczbami ujemnymi, jak i dodatnimi, a także mogą przyjmować wartość zero).

Formuły wartościujące w metodzie unitaryzacji zerowanej

Celem operacji normowania jest pozbawienie zmiennych mian oraz ujednolicenie ich przedziałów zmienności. Zmienne unormowane będziemy określać symbolem Z. Zbiór cech diagnostycznych podzielono na trzy podzbiory:

- S – stymulant,
- D – destymulant,
- N – nominant.

W MUZ najczęściej stosowane są następujące formuły normalizacyjne:

$$z_{ij} = \frac{x_{ij} - \min\limits_{i} x_{ij}}{\max\limits_{i} x_{ij} - \min\limits_{i} x_{ij}}, \qquad X_j \in S, \tag{4}$$

$$z_{ij} = \frac{\max\limits_{i} x_{ij} - x_{ij}}{\max\limits_{i} x_{ij} - \min\limits_{i} x_{ij}}, \qquad X_j \in D, \tag{5}$$

$$z_{ij} = \begin{cases} \dfrac{x_{ij} - \min\limits_{i} x_{ij}}{c_{0j} - \min\limits_{i} x_{ij}}, & \text{gdy } x_{ij} < c_{0j} \\ 1, & \text{gdy } x_{ij} = c_{0j} \\ \dfrac{x_{ij} - \max\limits_{i} x_{ij}}{c_{0j} - \max\limits_{i} x_{ij}}, & \text{gdy } x_{ij} > c_{0j} \end{cases}, \quad X_j \in N, \tag{6}$$

przy czym c_{0j} – nominalna wartość j-tej zmiennej,

oraz

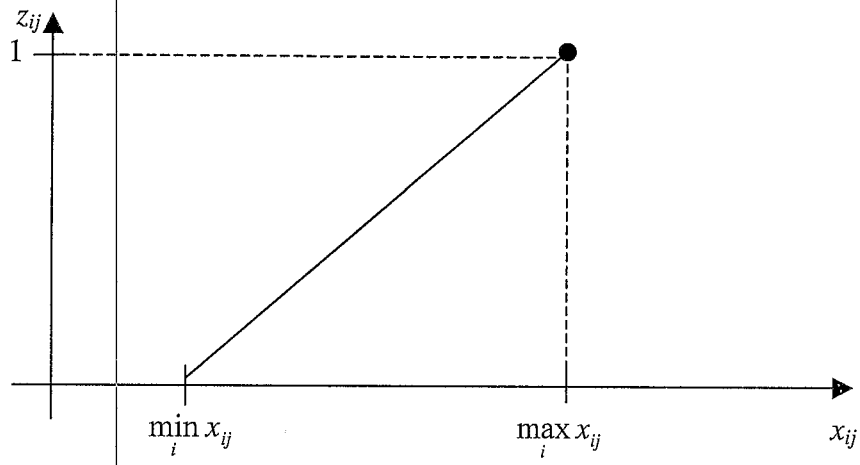
$$z_{ij} = \begin{cases} \dfrac{x_{ij} - \min\limits_{i} x_{ij}}{c_{1j} - \min\limits_{i} x_{ij}}, & \text{gdy } x_{ij} < c_{1j} \\ 1, & \text{gdy } c_{1j} \le x_{ij} \le c_{0j}, \\ \dfrac{x_{ij} - \max\limits_{i} x_{ij}}{c_{2j} - \max\limits_{i} x_{ij}}, & \text{gdy } x_{ij} > c_{2j} \end{cases} \qquad X_j \in N. \tag{7}$$

Formuła (7) jest przewidziana dla normowania nominant z określonym przedziałem wartości nominalnych $[c_{1j}, c_{2j}]$.

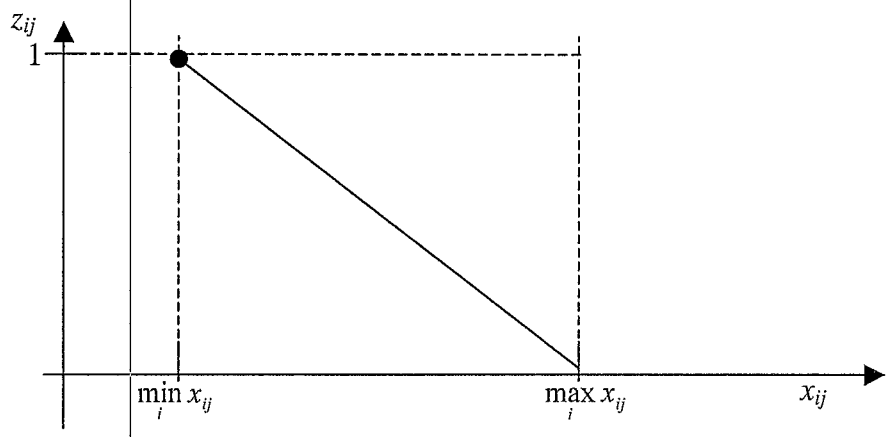
W każdej z przedstawionych formuł otrzymujemy:

$$z_{ij} \in [0,\ 1] \tag{8}$$

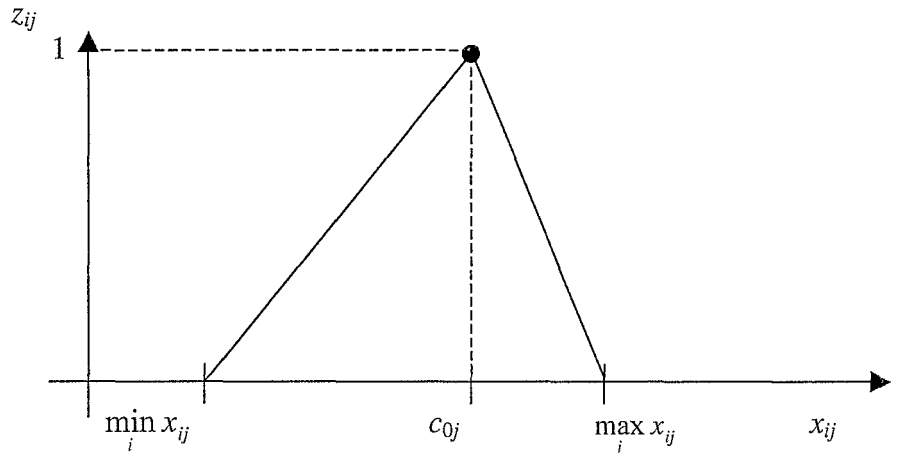
Działanie MUZ w swej klasycznej wersji opisane formułami (4)–(7) ilustrują rysunki 1–4.



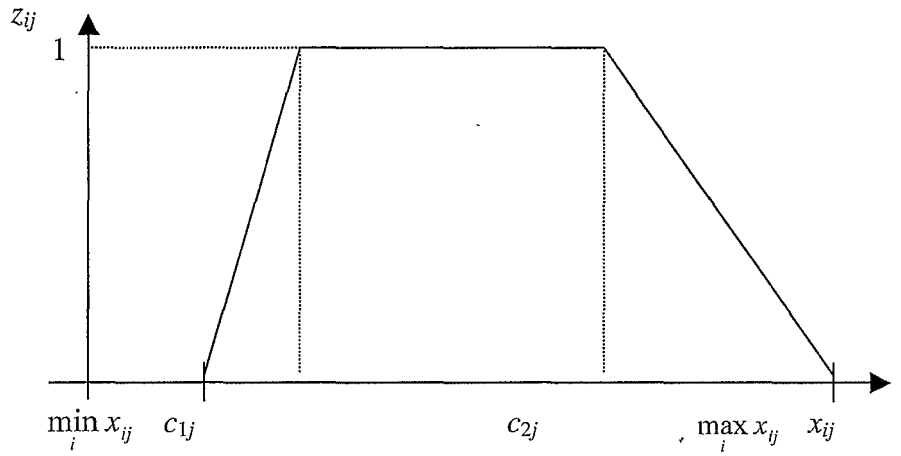
Rysunek 1
($X_j \in S$)



Rysunek 2
($X_j \in D$)



Rysunek 3
($X_j \in N$)



Rysunek 4
($X_j \in N$)

Postulaty stawiane metodom normującym cechy diagnostyczne

Normowanie cech diagnostycznych nie jest czynnością li tylko formalną, ma ono bowiem do spełnienia pewne cele natury praktycznej. Główne cele normowania według Borysa [2] można ująć następująco:

1) wprowadzenie addytywności w zbiorach wartości cech o różnych mianach;
2) określenie na zbiorze wartości cechy funkcji preferencyjnej; w zależności od rodzaju cechy funkcja preferencyjna jest transformacją stymulacyjną, destymulacyjną lub stymulacyjno-destymulacyjną (przypadek nominant).

Normowanie cech ma za zadanie umożliwienie realizacji szeroko zakrojonych badań porównawczych obiektów ze względu na poziom wielu zmiennych (cech) przyjętych jako kryteria oceny rozpatrywanego zjawiska złożonego. Aby zadanie to mogło być należycie wypełnione, nie należy traktować obojętnie własności charakteryzujących poszczególne formuły normalizacyjne. Jest rzeczą niezmiernie ważną, aby stosowane przez praktyków procedury normujące spełniały określone wymogi. Postulaty te można sformułować w kilku punktach:

1) pozbawienie mian (jednostek), w których są wyrażone cechy diagnostyczne;
2) sprowadzenie rzędu wielkości zmiennych diagnostycznych do stanu porównywalności, co oznacza wyrównanie zakresów zmienności cech, a w konsekwencji możliwość ich dodawania;
3) równość długości przedziałów zmienności wartości wszystkich cech unormowanych (stałość rozstępu z_{ij}) oraz równość dolnej i górnej granicy ich przedziału zmienności, w szczególności chodzi o przedział [0, 1];
4) możliwość normowania cech diagnostycznych przyjmujących wartości zarówno dodatnie, jak i ujemne lub tylko ujemne;
5) możliwość normowania cech przyjmujących wartość zero;
6) nieujemność wartości cech unormowanych;
7) istnienie prostych formuł – w ramach danej procedury normalizacyjnej – ujednolicających charakter zmiennych.

Warto podkreślić, że prawie wszystkie formuły normalizacyjne spełniają dwa pierwsze postulaty. Pozostałe postulaty uwzględniają tylko niektóre z formuł, i to nie zawsze wszystkie naraz. Można pokusić się o stwierdzenie, że metoda normalizacyjna, która spełnia wymienione postulaty, gwarantuje uniwersalne unormowanie wszystkich cech niezależnie od ich charakteru, rzędu wielkości czy też znaku.

Założono, iż w analizie porównawczej własności formuł normalizacyjnych brane są pod uwagę stymulanty jako zmienne najczęściej występujące w badaniach empirycznych. Ponadto, formuły dla destymulant i nominant są z reguły pokrewne konstrukcyjnie w stosunku do wzorów dla stymulant i mają te same bądź zbliżone do nich własności. W związku z tym wnioski płynące z analizy formuł przeznaczonych dla stymulant można uogólnić na wszystkie formuły.

Własności MUZ na tle pozostałych metod normowania

W tym podrozdziale starano się sprecyzować, a następnie poddać analizie własności dziesięciu wybranych formuł normujących cechy diagnostyczne (I––X) głównie pod kątem sformułowanych postulatów. Aby doprowadzić do zbiorczego ich porównania, zbudowano tabelę 1, w której w sposób dychotomiczny określono zgodność formuł normowania w stosunku do postulatów (1)––(7) oraz posiadania lub nieposiadania stałych wartości parametrów charakteryzujących zmienne unormowane. Z analizy zawartości tej tabeli wynika, iż żadna z formuł przekształceniowych nie ma samych plusów. Najwięcej jednak pozytywów wiąże się z formułą (V), właściwą metodzie unitaryzacji zerowanej. W dalszej kolejności plasują się formuły (I) i (IX). Gradacji tej nie należy kojarzyć jako jednoznacznego wskazania metod najlepszych, o wyborze bowiem metody w konkretnych zastosowaniach mogą decydować dodatkowe kryteria (tu nie uwzględnione) bądź też spełnienie lub niespełnienie jednego tylko z wymienionych postulatów.

Zauważmy na podstawie wyników zawartych w tabeli 1, że najtrudniejszym do spełnienia wymogiem w kontekście rozważanych formuł okazał się postulat (3). Postulat ten jest uszczegółowionym rozwinięciem postulatu (2), który tu traktujemy jako spełniony przez wszystkie formuły, bardziej w sensie intencjonalnym. Napotykamy dość wyraźne zróżnicowania wyników zarówno pod względem wartości rozstępu, jak i usytuowania dolnych i górnych granic przedziałów zmienności zmiennych unormowanych poszczególnymi metodami. Z wszystkich formuł branych pod uwagę tylko MUZ – formuła (V) – daje unormowania w stałym przedziale [0, 1]. Względnie ustabilizowane – wg przyjętych kryteriów – wyniki normowania uzyskano przy transformacjach opisanych wzorami (I), (IV), (IX) i (X).

Istnieje wiele trudności ze wskazaniem najlepszej metody odpowiadającej celom i zakresowi konkretnej analizy porównawczej. Na podstawie spostrzeżeń dokonanych w kontekście omawianych formuł transformacyjnych podejmuje-

my próbę sformułowania kilku wskazówek o charakterze aplikacyjnym.

Tabela 1
Zestawienie porównawcze formuł normujących cechy diagnostyczne

Formuła normująca	Postulaty stawiane formułom normującym							Stałość parametrów zmiennych unormowanych	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	$\overline{Z}_j$	$S^2(Z_j)$
(I)	+	+	–	+	+	–	–	+	+
(II)	+	+	–	+	+	–	–	–	+
(III)	+	+	–	+	+	–	+	–	–
(IV)	+	+	–	+	+	–	–	+	–
(V)	+	+	+	+	+	+	+	–	–
(VI)	+	+	–	–	–	+	+	–	–
(VII)	+	+	–	–	–	+	+	–	–
(VIII)	+	+	–	–	+	+	–	+	–
(IX)	+	+	–	–	+	+	+	+	–
(X)	+	+	–	–	+	+	+	–	–

Legenda:
(+) należy kojarzyć ze spełnieniem postulatu bądź ze stałością parametru charakteryzującego zmienną unormowaną,
(–) należy kojarzyć z niespełnieniem postulatu bądź z niestałością parametru charakteryzującego zmienną unormowaną.

Źródło: Opracowanie własne.

Duży wpływ na rezultaty porządkowania liniowego obiektów a tym samym na budowę ich rankingu ma wybór odpowiedniej formuły normującej. Przy czym w tym przypadku zaleca się wybór tych formuł, które dają stabilne bądź prawie stabilne przedziały zmienności zmiennych unormowanych, ze szczególnym wskazaniem MUZ.

Nieco innego wyboru metod można dokonać mając na celu modelowanie zjawisk złożonych. Tu często wykorzystuje się stałość parametrów charakteryzujących zmienne unormowane (por. E. Nowak [6], s. 80–81). W takich przypadkach można zalecać formuły (I), (VIII) lub (IX).

W przypadkach, w których sumowanie poszczególnych wartości realizacji zmiennych po obiektach przejawia sens, można zastosować formułę (IX). W wyniku tego zabiegu otrzymujemy struktury przestrzenne, które zawierają dodatkową informację o badanym zjawisku. Tego typu normowania można z powodzeniem

stosować w przestrzennych analizach produkcji przemysłowej, rolniczej itp.

Wreszcie kilka uwag o normowaniu zmiennych przyjmujących wartości ujemne. Należy zauważyć, iż stosunkowo rzadko napotykamy konieczność transformowania zmiennych przyjmujących wartości zarówno dodatnie, jak i ujemne. Niemniej istnieje kilka przypadków, w których konieczność ta wystąpi. Są to zwykle badania porównawcze dotyczące kondycji finansowej firm, banków i innych instytucji. W badaniach tych nie sposób pominąć kategorii określonej mianem wynik finansowy, który może przybierać wartości zarówno dodatnie, jak i zero czy też wartości ujemne. Musimy zatem dobrać taką metodę normowania, która transformuje zmienne diagnostyczne o wszystkich możliwych wartościach ($X_j \in R$). W tej sytuacji mogą być wykorzystane formuły (I), (IV) i (V).

Literatura

1. BARTOSIEWICZ S., Propozycja metody tworzenia zmiennych syntetycznych, *Prace Naukowe Akademii Ekonomicznej we Wrocławiu*, 1976, nr 84.
2. BORYS T., Metody normowania cech w statystycznych badaniach porównawczych, *Przegląd Statystyczny* 1978, z. 2.
3. CIEŚLAK M., *Taksonomiczna procedura programowania rozwoju gospodarczego i określania zapotrzebowania na kadry kwalifikowane*, PWN, Warszawa 1976.
4. GRABIŃSKI T., Wielowymiarowa analiza porównawcza w badaniach dynamiki zjawisk gospodarczych, *Zeszyty Naukowe Akademii Ekonomicznej w Krakowie*, seria specjalna: *Monografie*, nr 61, Kraków 1984.
5. KUKUŁA K., *Metoda unitaryzacji zerowanej*, Wydawnictwo Naukowe PWN Warszawa 2000.
6. NOWAK E., Propozycja prostej metody konstruowania miernika rozwoju i jego wykorzystania do badań regresyjnych, *Przegląd Statystyczny* 1979, z. 1–2.
7. STRAHL D., Propozycja konstrukcji miary syntetycznej, *Przegląd Statystyczny* 1978, z. 2.

Properties of Zero Unitarization Method

Abstract

The paper presents properties of zero unitarization method in relation with some other normalisation methods.

Formulas for normalising stimulants, destimulants and nominants are shown and compared. Special attention is given to the problem of normalisation of nominants.

In the concluding part the method discussed in the paper is proven to be universal and easy in many applications.

Bolesław Borkowski, Hanna Dudek
Katedra Ekonometrii i Informatyki SGGW

Analiza zjawisk ekonomicznych za pomocą modeli wielorównaniowych

Wstęp

W badaniach ekonomicznych bardzo często chcemy określić ilościowe związki pomiędzy efektem (produkcją, zyskiem, dochodem itp.) a nakładem lub nakładami. Do tego celu wykorzystywane są modele ilościowych zależności między zjawiskami ekonomicznymi wyrażonymi w postaci równań. Statystyczne metody ekonometrii, szczególnie jednowymiarowe równania opisowe, które ujmują ilościowe zależności między różnymi wielkościami ekonomicznymi i pozaekonomicznymi, są często i wielostronnie wykorzystywane w badaniach ekonomicznych. W przypadku zależności między zmienną objaśnianą a zespołem zmiennych objaśniających możemy wprowadzić zespół zmiennych, o których wiemy, że łączy je zależność funkcyjna o postaci:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + \varepsilon$$

gdzie: Y jest zmienną objaśnianą, X_j – zmiennymi objaśniającymi, ε – składnikiem losowym, a β_j, dla j = 1, 2, 3, ..., k, parametrami, które na ogół są nieznane (wartości te są szacowane na podstawie *n*-elementowej próby (Y_i, X_{1i}, X_{2i}, ..., X_{ki}) dla i = 1, 2, 3, ..., n).

Dotychczasowe badania ekonometryczne koncentrowały się głównie na metodach doboru zmiennych objaśniających do jednorównaniowego modelu ekonometrycznego oraz postaci analitycznej funkcji aproksymacyjnej. W literaturze naukowej proponowane są różne kryteria wyboru postaci analitycznej funkcji aproksymacyjnej oraz kilka sposobów szacowania nieznanych parametrów wybranej funkcji. Dotychczas w teorii statystyki i ekonometrii nie opracowano jednoznacznych, możliwych do zastosowania metod wyboru postaci analitycznej funkcji. Stosowane kryteria są obarczone czynnikiem subiektywnym (por. [1]). Dlatego też najczęściej do badań zjawisk ekonomicznych stosowane są liniowe modele jednorównaniowe.

Jednorównaniowe modele ekonometryczne są dalece niewystarczające do odwzorowania bardziej złożonych procesów ekonomicznych. Modele jednorównaniowe mają duże zastosowanie tylko wówczas, gdy chcemy opisać jedną prawidłowość ekonomiczną w oderwaniu od innych. Takie podejście napotyka na dużą barierę, ponieważ nawet najlepiej wyspecyfikowane jednorównaniowe modele nie mogą w sposób poprawny i wyczerpujący opisać rzeczywistości gospodarczej. Zjawiska ekonomiczne i społeczne charakteryzują związki wzajemne (sprzężone), które zachodzą pomiędzy zmiennymi. Związki te są w ogólności stochastyczne, dynamiczne i – co należy podkreślić – bardzo często jednoczesne. Analiza tych zjawisk musi więc uwzględniać wzajemne powiązania pomiędzy zmiennymi a więc nie może być opisana jednorównaniowym modelem, lecz wielorównaniowym. Modele wielorównaniowe odznaczają się tym, że nie nakłada się żadnych ograniczeń na rodzaj i liczbę powiązań pomiędzy zmiennymi endogenicznymi modelu. W modelach tych dopuszczalne jest zjawisko współzależności dwóch lub więcej zmiennych endogenicznych w tym samym czasie. Zjawiska ekonomiczne mają tu więc charakter obustronny lub wielostronny.

W literaturze przedmiotu wielorównaniowe modele ekonometryczne znane są od wielu lat. Przykładem takich modeli są modele Kleina (por. [2]), skonstruowane w latach czterdziestych, odnoszące się do opisu gospodarki USA. Zasadniczym problemem wykorzystania modeli wielorównaniowych w praktyce, oprócz ich identyfikacji (por. [7]), były trudności oszacowania ich ocen parametrów. Powszechnie stosowana do szacowania ocen parametrów równania jednorównaniowego metoda najmniejszych kwadratów w przypadku modeli wielorównaniowych o równaniach współzależnych nie może być wykorzystana. Estymatory klasycznej metody najmniejszych kwadratów tracą cenne własności zgodności i efektywności. Obecnie szybki rozwój techniki komputerowej oraz użytkowego oprogramowania pozwala na znajdowanie ocen parametrów modeli wielorównaniowych za pomocą specjalnych metod.

W niniejszej pracy podjęliśmy pierwszą próbę aplikacji liniowego modelu wielorównaniowego do badania wzajemnych związków pomiędzy efektami a czynnikami produkcji w gospodarstwach rolniczych.

Metody analizy

Wielorównaniowy liniowy model ekonometryczny można zapisać w następującej, ogólnej postaci (por. [4]):

$$BY + \Gamma Z = \varepsilon$$

gdzie:

Y – wektor (m × 1) zmiennych endogenicznych bez opóźnień czasowych,
B – macierz (m × m) parametrów przy zmiennych endogenicznych bez opóźnień czasowych,
Z – wektor (k × 1) zmiennych z góry ustalonych,
Γ – macierz (m × k) parametrów przy zmiennych z góry ustalonych,
ε – wektor (m × 1) odchyleń losowych.

Do budowy liniowego modelu wielowymiarowego o równaniach współzależnych wykorzystaliśmy dane empiryczne pochodzące z 1996 roku z 284 gospodarstw rolniczych o powierzchni od 5 do 25 ha UR. Dane te zostały zgromadzone na potrzeby badań realizowanych w Katedrze Ekonomiki i Organizacji Gospodarstw Rolniczych SGGW w ramach grantu KBN nr 5 PO6B037 nt. „Przestrzenne zróżnicowanie technologii produkcji roślinnej w Polsce i jego skutki". Każde gospodarstwo, po dokonaniu wstępnej eliminacji zmiennych quasi-stałych, opisane było 26 cechami (por. załącznik 1), których wartości zostały przeliczone na 1 ha UR. Spośród tych cech dokonaliśmy doboru zmiennych endogenicznych i egzogenicznych. Do potencjalnego zbioru zmiennych łącznie współzależnych (endogenicznych nieopóźnionych) zakwalifikowaliśmy następujące zmienne:

- ✓ Produkcja roślinna w zł/ha.
- ✓ Produkcja zwierzęca w zł/ha.
- ✓ Produkcja końcowa w zł/ha.
- ✓ SD na 1 ha.
- ✓ Plon przeliczeniowy w dt/ha.
- ✓ Koszty działalności rolniczej w zł/ha.

Pozostałe zmienne wymienione w załączniku traktowane były jako zmienne z góry ustalone. Po wstępnej analizie współczynników korelacji pomiędzy wszystkimi rozważanymi zmiennymi za zmienne endogeniczne przyjęliśmy produkcję roślinną w zł/ha, liczbę sztuk dużych w przeliczeniu na hektar oraz plon przeliczeniowy w dt/h.

Dobór zmiennych objaśniających do poszczególnych równań modeli współzależnych jest zagadnieniem bardziej skomplikowanym niż w przypadku modeli jednowymiarowych czy też modeli wielowymiarowych prostych i rekurencyjnych. Należy uwzględniać tu wielokierunkowe powiązania między zmiennymi endogenicznymi bez opóźnień czasowych oraz zapewniać identyfikowalność równań modelu. W literaturze przedmiotu proponowane są różne rozwiązania. W pracy doboru zmiennych objaśniających dokonaliśmy za po-

mocą procedury dwustopniowej (por. [5]). W pierwszym etapie wybraliśmy zmienne objaśniające danego równania jedynie spośród wymienionych zmiennych endogenicznych bez opóźnień czasowych, odgrywających w tym równaniu rolę potencjalnych zmiennych objaśniających. Jako kryterium doboru zastosowaliśmy maksymalną integralną pojemność informacyjną tych zmiennych (metoda Hellwiga). W drugim etapie przeprowadziliśmy dodatkowy dobór zmiennych objaśniających spośród pozostałych zmiennych z góry ustalonych, grających rolę potencjalnych zmiennych objaśniających. Rozpatrywaliśmy wszystkie możliwe kombinacje potencjalnych zmiennych z góry ustalonych z wszystkimi zmiennymi współzależnymi, wybranymi w pierwszym etapie. Wykorzystaliśmy takie samo kryterium doboru jak w etapie pierwszym. Dodatkowo rozpatrywaliśmy konieczny warunek identyfikowalności (por. [2], [6]). W ten sposób skonstruowaliśmy model wielorównaniowy o trzech równaniach współzależnych z następującymi zmiennymi:

- ✓ W równaniu pierwszym produkcja roślinna jako zmienna endogeniczna bez opóźnień czasowych objaśniana jest następującymi zmiennymi:
  - udziałem gleb dobrych w % – zmienna egzogeniczna,
  - wartością maszyn w zł/ha – zmienna egzogeniczna,
  - plonem przeliczeniowym w dt/ha[1] – zmienna endogeniczna.

- ✓ W równaniu drugim SD na ha jako zmienna endogeniczna bez opóźnień czasowych objaśniana jest następującymi zmiennymi:
  - pracą w osobach na ha – zmienna egzogeniczna,
  - powierzchnią TUZ w ha – zmienna egzogeniczna,
  - kosztem energii i paliw w zł na ha – zmienna egzogeniczna,
  - produkcją roślinną w zł/ha – zmienna endogeniczna.

- ✓ W równaniu trzecim plon przeliczeniowy jako zmienna endogeniczna bez opóźnień czasowych objaśniana jest następującymi zmiennymi:
  - NPK w dt na ha – zmienna egzogeniczna,
  - kosztem materiału siewnego w zł/ha – zmienna egzogeniczna,
  - kosztem pestycydów w zł /ha – zmienna egzogeniczna,
  - sztuki duże (SD) na ha – zmienna endogeniczna.

---

[1]W celu obliczenia plonu przeliczeniowego w dt/ha zsumowano wielkość produkcji wszystkich zbóż, wielkość produkcji ziemniaków podzieloną przez 7 oraz wielkość produkcji buraków podzieloną przez 12. Suma tak obliczonej wielkości produkcji została podzielona przez powierzchnię przeznaczoną pod uprawę zbóż, ziemniaków i buraków.

Powyższy model jest więc modelem wielorównaniowym o równaniach współzależnych oraz jest on niejednoznacznie identyfikowalny.

Następnym trudnym problemem w modelach wielowymiarowych jest estymacja parametrów.

Metody estymacji parametrów modeli wielorównaniowych można podzielić na:

- procedury estymacji poszczególnych równań oddzielnie,
- metody łącznej estymacji wszystkich równań modelu.

W metodach estymacji pojedynczych równań ignorowane są związki korelacyjne pomiędzy składnikami losowymi pochodzącymi z różnych równań. Gdy składniki losowe pochodzące z różnych równań są skorelowane, bardziej efektywne są estymatory metod łącznej estymacji wszystkich równań modelu.

W przypadku modeli o równaniach współzależnych do najczęściej stosowanych metod estymacji należą potrójna metoda najmniejszych kwadratów (3MNK) oraz metoda największej wiarygodności z pełną informacją (NWPI) (por. [7]).

Wymienione metody są ekwiwalentne w przypadku, gdy każde równanie strukturalne jest jednoznacznie identyfikowalne. W naszym przypadku model jest niejednoznacznie identyfikowalny.

W celu porównania wyników estymacji do oszacowania ocen parametrów skonstruowanego modelu wielorównaniowego zastosowaliśmy dwie wymienione metody. W obu przypadkach zbadaliśmy jakość ocen parametrów strukturalnych oraz stopień zgodności uzyskanych modeli z danymi empirycznymi. Uwzględniliśmy standardowe błędy szacunku parametrów, odchylenia standardowe reszt i współczynniki korelacji wielorakiej dla poszczególnych równań modelu oraz wartość funkcji wiarygodności[2].

Przy estymacji ocen parametrów modelu potrójną metodą najmniejszych kwadratów logarytm funkcji wiarygodności wynosił –2011,2 (po zlogarytmowaniu zmiennych, log-likehood = 506,1), natomiast przy zastosowaniu metody największej wiarygodności z pełną informacją logarytm funkcji wiarygodności wynosił –2008,0 (po zlogarytmowaniu zmiennych, log-likehood = 533,4). W obu przypadkach standardowe błędy szacunku parametrów nie przekraczały 50% wartości parametrów. Z uwagi na nieco lepsze wyniki estymacji metodą największej wiarygodności z pełną informacją do badań przyjęliśmy model oszacowany tą metodą.

---

[2]Z przyczyn numerycznych nie oblicza się wartości funkcji wiarygodności, lecz jej logarytm naturalny (log-likehood function). Wyższe wartości logarytmu wiarygodności świadczą o lepszym dopasowaniu modelu do danych empirycznych.

Następnie przeprowadziliśmy badania w celu stwierdzenia, czy składniki losowe pochodzące z trzech równań naszego modelu są skorelowane. Wielkości współczynników korelacji składników losowych uzyskanych przy estymacji ocen parametrów modelu metodą największej wiarygodności z pełną informacją przedstawiają dane w tabeli 1.

Tabela 1
Wielkości współczynników korelacji pomiędzy składnikami losowymi

Wyszczególnienie	Produkcja roślinna w zł/ha	SD na 1 ha	Plon przeliczeniowy
Produkcja roślinna w zł/ha	1,0000		
SD na 1 ha	–0,0849	1,0000	
Plon przeliczeniowy	–0,4573	0,8618	1,0000

Źródło: Obliczenia własne na podstawie danych empirycznych wykonane pakietem PcGive 9.2.

Wiadomo, że estymatory NWPI są asymptotycznie bardziej efektywne w stosunku do estymatorów metod, za pomocą których szacuje się parametry każdego równania z osobna, gdy składniki losowe pochodzące z różnych równań są skorelowane (por. [7]). W naszym przypadku korelacja pomiędzy składnikami losowymi jest znaczna.

W celu porównania dokonaliśmy także oszacowania ocen parametrów jednowymiarowego modelu ekonometrycznego, przyjmując jako zmienną objaśnianą produkcję roślinną w zł na ha. Na podstawie tego samego materiału empirycznego dokonaliśmy doboru zmiennych objaśniających do jednowymiarowego modelu metodą regresji krokowej (forward selection i backward selection) a oceny parametrów jednowymiarowego modelu oszacowaliśmy MNK.

Wyniki badań

Model wielorównaniowy po oszacowaniu ocen parametrów strukturalnych ma postać:

$$\hat{Y}_1 = \underset{(9,83)}{66,7\,Y_3} + \underset{(193,98)}{1001,8\,X_1} + \underset{(0,012)}{0,036\,X_2} - \underset{(275,1)}{1802,7}$$

$$\hat{Y}_2 = -0{,}000136\,Y_1 + 0{,}004818\,X_3 + 0{,}01393\,X_4 + 0{,}0000975\,X_5 + 0{,}6587$$
$$(0{,}0000222) \quad (0{,}00052) \quad (0{,}0034) \quad (0{,}000036) \quad (0{,}039)$$

$$\hat{Y}_3 = -34{,}97\,Y_2 + 0{,}02377\,X_6 + 0{,}01232\,X_7 + 0{,}01768\,X_8 + 49{,}99$$
$$(6{,}42) \quad (0{,}0052) \quad (0{,}0036) \quad (0{,}0062) \quad (4{,}56)$$

gdzie:
Y_1 – wartość produkcji roślinnej w zł na 1 ha UR,
Y_2 – liczba SD na 1 ha UR,
Y_3 – plon przeliczeniowy w dt/ha,
X_1 – udział gleb dobrych w %,
X_2 – wartość maszyn w zł na 1 ha UR,
X_3 – praca w osobach na 1 ha UR,
X_4 – powierzchnia TUZ w ha,
X_5 – koszty energii i paliw w zł na 1 ha UR,
X_6 – NPK w kg na 1 ha UR,
X_7 – koszty materiału siewnego w zł na 1 ha UR,
X_8 – koszty pestycydów w zł na 1 ha UR,
( ) – wartości standardowych błędów szacunku parametrów strukturalnych.

Ocenę dopasowania modelu wielowymiarowego do danych empirycznych przeprowadziliśmy za pomocą wielkości odchylenia standardowego (Se) i współczynnika korelacji wielorakiej (R) poszczególnych równań. Wielkości tych estymatorów zamieszczone są w tabeli 2.

Tabela 2
Wielkości S_e i R

Równanie	Wielkości S_e	Wielkości R
Pierwsze	725,73	0,70
Drugie	0,28	0,69
Trzecie	12,42	0,77

Otrzymane wielkości wskazują na niezłe dopasowanie poszczególnych równań modelu wielowymiarowego do danych empirycznych. Dodatkowo przeprowadzona weryfikacja istotności parametrów strukturalnych modelu wielorównaniowego wykazała ich istotność na poziomie $\alpha = 0{,}01$. Natomiast badania jednorodności wariancji składników losowych wykazały, że składniki losowe występujące w poszczególnych równaniach są sferyczne.

W oszacowanym modelu wielorównaniowym występują wielostronne zależności pomiędzy zmiennymi. Uwzględnione zmienne odgrywają różną rolę w poszczególnych równaniach. Zmienna objaśniana Y1 – wartość produkcji roślinnej w zł/ha w równaniu pierwszym jest zmienną objaśniającą w równaniu drugim. Natomiast zmienna objaśniana Y2 – liczba SD na 1 ha UR w równaniu drugim jest zmienną objaśniającą w równaniu trzecim itp. Gdy te same zmienne odgrywają różną rolę w poszczególnych równaniach, wówczas model taki należy do klasy modeli wielorównaniowych o równaniach współzależnych. Oszacowane oceny parametrów takiego modelu należy interpretować łącznie, a nie rozdzielnie, jak w przypadku modeli jednorównaniowych (por. [8]). Produkcja roślinna w badanych gospodarstwach była determinowana przez wyżej wymieniony zespół zmiennych, uwzględnionych w trzech równaniach modelu wielorównaniowego. Widoczne są wzajemne związki występujące pomiędzy wielkością produkcji roślinnej a technologią uprawy oraz wielkością produkcji zwierzęcej. Wzrost udziału gleb dobrych o 1% w badanych gospodarstwach wpływał na zwiększenie produkcji roślinnej średnio o ponad 1000 zł/ha przy średnim poziomie pozostałych zmiennych. Natomiast wzrost wyposażenia badanych gospodarstw w maszyny i urządzenia o 10 tys. zł wpływał na zwiększenie produkcji roślinnej średnio o około 400 zł/ha przy średnim poziomie pozostałych zmiennych uwzględnionych w modelu wielorównaniowym. Zaobserwowano także wpływ wielkość plonu przeliczeniowego na produkcję roślinną w badanych gospodarstwach. Wraz ze wzrostem plonu przeliczeniowego o 1 dt/ha wartość produkcji roślinnej zwiększała się średnio o blisko 67 zł/ha (w badanych okresie średnia cena 1 dt zbóż wynosiła 300 zł, 1 dt ziemniaków 30 zł, a buraków 10 zł). Wielkość plonu przeliczeniowego zależna była od stosowanej technologii uprawy (wielkości NPK, kosztu materiału siewnego, kosztu stosowanych pestycydów) oraz od wielkości produkcji zwierzęcej wyrażonej w modelu przez obsadę inwentarza żywego na 100 ha UR. Zależność pomiędzy liczbą inwentarza żywego w badanych gospodarstwach a wielkością plonu była ujemna. Jest to bardzo interesujący rezultat. Można więc sądzić, że w badanej populacji gospodarstwa o przewadze produkcji zwierzęcej mniejszą wagę przywiązywały do wielkości towarowej produkcji roślinnej. Na ujemny efekt wzrostu pogłowia zwierząt mierzony wielkością plonu przeliczeniowego wpłynęły głównie czynniki wytwórcze, które determinowały liczbę SD na 1 ha UR w badanych gospodarstwach, tj. powierzchnia TUZ, koszt energii i paliw, praca ludzka i wielkość produkcji roślinnej. Wraz ze wzrostem produkcji roślinnej o 10 tys. zł wielkość obsady zwierząt w badanych gospodarstwach zmniejszała się średnio o ponad 1 SD na 1 ha UR.

Oszacowane oceny parametrów modelu wielorównaniowego wskazują wyraźnie na wielokierunkowe zależności pomiędzy czynnikami wytwórczymi w gospodarstwach rolniczych. Związki te są szczególnie wyraźne pomiędzy produkcją roślinną a produkcją zwierzęcą, które w badanych gospodarstwach były obustronne.

Oszacowane oceny parametrów modelu jednorównaniowy klasyczną MNK, w którym zmienną objaśnianą była produkcja roślinna w zł/ha, mają następujące wielkości:

$$\hat{Y}_1 = 1{,}8016\,X_1 + 30{,}83\,X_2 - 27{,}33\,X_3 - 1016{,}5\,X_4 + 879{,}5\,X_5 + 0{,}0629\,X_6$$
$$(0{,}6133) \quad (4{,}78) \quad (6{,}16) \quad (110{,}9) \quad (163{,}6) \quad (0{,}015)$$

$$R^2 = 79{,}18$$

gdzie:
Y_1 – wartość produkcji roślinnej w zł na 1 ha UR,
X_1 – NPK w kg na 1 ha UR,
X_2 – plon przeliczeniowy w dt/ha,
X_3 – powierzchnia UR w ha,
X_4 – liczba SD na 1 ha UR,
X_5 – udział gleb dobrych w %,
X_6 – wartość maszyn w zł na 1 ha UR,
( ) – wartości standardowych błędów szacunku parametrów strukturalnych.

Powyższy model został oszacowany na podstawie tej samej próby. Doboru zmiennych do modelu dokonaliśmy metodą regresji krokowej (forward selection). Liczba zmiennych objaśniających produkcję roślinną jest tu znacznie mniejsza niż w modelu wielorównaniowym. Oszacowane oceny parametrów wskazują tu na jednokierunkowe powiązania czynników produkcji ze zmienną objaśnianą. Zwiększenie wielkości stada zwierząt o 1 SD/ha powoduje spadek produkcji roślinnej średnio o ponad 1000 zł/ha. Zachowany jest tu taki sam kierunek zależności pomiędzy wielkością produkcji roślinnej i zwierzęcej, ale wartość estymatora jest znacznie większa niż w modelu wielorównaniowym. Wielkość tego estymatora zawiera prawdopodobnie wartość współzmienności tej zmiennej z innymi zmiennymi nie uwzględnionymi w modelu jednorównaniowym. Nie jesteśmy w stanie określić tej wielkości na podstawie oszacowa-

nego modelu jednowymiarowego. Natomiast taką analizę wzajemnych powiązań mogliśmy przeprowadzić na podstawie estymatorów modelu wielorównaniowego. Trudno także jednoznacznie zinterpretować bezpośredni wpływ wielkości nawożenia na wartość produkcji roślinnej. Raczej wpływ tej zmiennej jest przez wielkość plonu, która jest determinowana wielkością nawożenia mineralnego, itp. Takiej analizy wzajemnych powiązań nie możemy przeprowadzić na podstawie jednorównaniowego modelu ekonometrycznego.

Podsumowanie i wnioski

Przeprowadzone badania wykazały, że modele wielorównaniowe mogą być bardzo przydatne do analizy wzajemnych powiązań czynników produkcji w gospodarstwach rolniczych. Badania wykazały, że związki przyczynowe nie są tu jednostronne i nie przebiegają jednokierunkowo. Oszacowane oceny parametrów modelu wielorównaniowego wskazały wyraźnie na wielokierunkowe zależności pomiędzy czynnikami wytwórczymi w gospodarstwach rolniczych. Związki te były szczególnie wyraźne pomiędzy produkcją roślinną o produkcją zwierzęcą i w badanych gospodarstwach były obustronne. Szczególnie zaobserwowano wzajemne związki występujące pomiędzy wielkością produkcji roślinnej a technologią uprawy i wielkością produkcji zwierzęcej.

Celem badań było także wskazanie odpowiednich metod doboru zmiennych do modeli wielorównaniowych oraz metod oszacowania ocen parametrów modelu. W zależności od klasy modeli wielorównaniowych można stosować różne metody doboru zmiennych. W przypadku modeli wielowymiarowych o równaniach współzależnych doboru zmiennych objaśniających dobre rezultaty daje metoda doboru zmiennych według procedury dwustopniowej. Równorzędnymi metodami estymacji okazały się być potrójna metoda najmniejszych kwadratów oraz metoda największej wiarygodności z pełną informacją.

Literatura

1. BORKOWSKI B. SZCZESNY W.: Zastosowanie funkcji Cobb-Douglasa w badaniach ekonomiczno-rolniczych, *ZER* nr 4/1985, s. 49–61.
2. GREENE W. H.: *Econometric Analysis*, Prentice Hall, New York 2000.
3. KLEIN L.R.: *Wstęp do ekonometrii*, PWE, Warszawa 1965.
4. NOWAK E.: *Zarys metod ekonometrii. Zbiór zadań*, PWN, Warszawa 1994.

5. NOWAK E.: *Problemy doboru zmiennych do modelu ekonometrycznego*, PWN, Warszawa 1984.
6. RAMANATHAN R.: *Introductory Econometrics with Applications*, The Dryden Press, Fort Worth 1998.
7. WELFE A.: *Ekonometria*, PWE, Warszawa 1998.
8. WOŚ A.: O niektórych postaciach funkcji produkcji w rolnictwie, *Ekonomista* nr 4, 1964.

Załącznik 1

Lista zmiennych uwzględnionych przy budowie modelu:

1. Wartość produkcji roślinnej [zł na ha UR].
2. Wartość produkcji zwierzęcej [zł na ha UR].
3. Wartość produkcji końcowej [zł na ha UR].
4. Powierzchnia TUZ [ha].
5. Powierzchnia GO [ha].
6. Powierzchnia UR [ha].
7. Udział gleb dobrych [%].
8. Praca [osoby na 1 ha].
9. Liczba SD w przeliczeniu na ha.
10. Wartość budynków [zł na ha].
11. Wartość maszyn [zł na ha].
12. Kubatura budynków [m^3].
13. Liczba maszyn.
14. Zapasy produktów roślinnych na początku roku [zł na ha].
15. Plon przeliczeniowy [dt na ha].
16. NPK [kg na ha].
17. Koszty materiału siewnego [zł na ha].
18. Koszty pestycydów [zł na ha].
19. Koszty produkcji roślinnej [zł na ha].
20. Koszty usług weterynaryjnych [zł na ha].
21. Koszty zakupu pasz [zł na ha].
22. Koszty zakupu zwierząt [zł na ha].
23. Koszty produkcji zwierzęcej [zł na ha].
24. Koszty energii i paliw [zł na ha].
25. Koszty pośrednie [zł na ha].
26. Koszty działalności rolniczej [zł na ha].

The analysis of economic fenomena by using simultaneous equation models

Abstract

Widely applied single-equation econometric model deals with a single dependent variable. In more complex systems such model is not an adequate tool. Economicfenomena are characterized by interactions. In such cases simultaneous equation model is more appropriate. In simultaneous equation models there are no limits on kind and number of connections between endogenous variables. They can be applied everywhere where relationship between variables is not one-sided. The attempt of application of linear simultaneous model to analysis of interdependence between inputs and outputs in farm has been undertaken in this paper. There some data from 284 farms have been used in this research. For choosing explanatory variables two-stages procedure based on Hellwig's methods has been applied. In order to estimate the parameters joint estimation of entire system of equation – there-state least squares and full-information maximum likelihood methoods have been used.

The endogenous variables of that model are: value of vegetable production, crop and livestock.

Joanna Kisielińska
Katedra Ekonometrii i Informatyki SGGW

Elementy teorii optymalizacji oraz metody rozwiązywania zadań optymalizacji nieliniowej

Wprowadzenie

Wiele zagadnień z zakresu ekonomii można przedstawić w postaci zadań optymalizacyjnych. Przykładem mogą być problemy maksymalizacji dochodów w warunkach strategii krótko- i długookresowej [4]. W praktyce powszechnie stosowane są liniowe modele optymalizacyjne, rozwiązywane za pomocą algorytmu simpleks. Znacznie rzadziej natomiast wykorzystywane są modele nieliniowe. Wynika to nie tyle z mniejszej przydatności tych modeli, co z niedostatecznej ich znajomości. Modele nieliniowe są trudniejsze do konstruowania oraz rozwiązywania. W niniejszym artykule przedstawione zostaną pokrótce teoretyczne podstawy oraz algorytmy optymalizacji nieliniowej.

Elementy teorii optymalizacji

Findeisen, Szymanowski i Wierzbicki [2] podają ogólną definicję optymalizacji: „Optymalizacja jest to postępowanie, polegające na wyborze elementu z danego zbioru w oparciu o relacje, ustalające pewien porządek w tym zbiorze". Zbiór ten nazywany jest zbiorem rozwiązań dopuszczalnych. Porządek w zbiorze rozwiązań dopuszczalnych może ustalać funkcja rzeczywista, zwana wskaźnikiem jakości lub funkcją celu. Funkcja ta jest określona dla zmiennych zwanych decyzyjnymi. Zadanie optymalizacji polega na takim doborze zmiennych decyzyjnych, aby funkcja celu osiągała wartość maksymalną lub minimalną.

W literaturze teoria optymalizacji formułowana jest zwykle dla zadania minimalizacji i dlatego też tak zostanie przedstawiona w dalszej części tej pracy. Poszukiwanie maksimum funkcji f(X) zawsze może być sprowadzone do poszukiwania minimum, ponieważ maksymalizacja f(X) jest równoważna minimalizacji – f(X).

Zadanie optymalizacji lub zadanie programowania matematycznego polega więc na poszukiwaniu:

$$\min f(\mathbf{X}) \tag{1}$$

$\mathbf{X} \in Z_R$, gdzie:

f – funkcja celu. Jest to funkcja n-zmiennych, przekształcająca n-wymiarową przestrzeń rzeczywistą R^n w zbiór liczb rzeczywistych R^1.

$\mathbf{X}$ – jest n-wymiarowym wektorem zmiennych decyzyjnych, czyli $\mathbf{X} \in R^n$. Z_R – jest zbiorem rozwiązań dopuszczalnych.

Jeżeli nie ma żadnych ograniczeń narzuconych na wybór rozwiązania (czyli, gdy $Z_R = R^n$), mówimy o zadaniu programowania matematycznego bez ograniczeń zapisywanym jako:

$$\min f(\mathbf{X}) \tag{2}$$

$\mathbf{X} \in R^n$

Jeśli natomiast zmienne decyzyjne muszą spełniać dodatkowe kryteria, tzw. kryteria dopuszczalności (czyli, gdy $Z_R \subset R^n$), zadanie optymalizacji nazywane jest zadaniem programowania matematycznego z ograniczeniami. Ograniczenia formułowane są w postaci równań lub nierówności. Znalezienie rozwiązania optymalnego polega wówczas na takim doborze wartości zmiennych decyzyjnych, aby spełnione były ograniczenia, a funkcja celu osiągała wartość maksymalną lub minimalną. Zadanie optymalizacji z ograniczeniami można zapisać następująco:

$$\min f(\mathbf{X}) \tag{3}$$

$\mathbf{X} \in Z_R = \{\mathbf{X}: g_i(\mathbf{X}) \leq 0, i = 1, ..., m\}$

gdzie:

$g_i: R^n \rightarrow R^1$, dla i = 1, ..., m – funkcje ograniczeń.

Wiele problemów ekonomicznych można sprowadzić do zagadnienia poszukiwania wartości ekstremalnych pewnych wielkości. Modele optymalizacyjne bez ograniczeń reprezentują problem maksymalizacji dochodu przedsiębiorstwa w warunkach strategii długookresowej, natomiast modele z ograniczeniami – w warunkach strategii krótkookresowej [4].

Dla zadania optymalizacji bez ograniczeń i z ograniczeniami formułowane są warunki konieczne oraz konieczne i wystarczające istnienia rozwiązania. W przypadku optymalizacji bez ograniczeń są to warunki istnienia minimum globalnego funkcji n zmiennych.

Warunkiem koniecznym istnienia minimum funkcji f: $R^n \rightarrow R^1$ w punkcie $\mathbf{X}^*$ jest to, aby jej gradient (wektor pochodnych cząstkowych) w tym punkcie był równy zeru.

Warunkiem koniecznym i wystarczającym istnienia dokładnie jednego punktu $\mathbf{X}^*$, w którym funkcja f: $R^n \rightarrow R^1$ osiąga minimum globalne jest:

1) spełnienie warunku koniecznego;
2) dodatnia półokreśloność hesjanu (macierzy drugich pochodnych) funkcji f, dla każdego $\mathbf{X}$.

Warunek konieczny (zerowanie gradientu) pozwala znaleźć punkt optymalny jedynie dla prostych zadań. Problem sprowadza się do rozwiązania układu n równań. Jeśli funkcja f jest kwadratowa, układ ten jest układem liniowym i wówczas jego rozwiązanie jest możliwe. W ogólnym przypadku jest to jednak układ nieliniowy, a wobec tego zwykle nie da się go rozwiązać analitycznie.

Jeśli nie można znaleźć rozwiązania zadania optymalizacji bez ograniczeń w sposób analityczny, należy posłużyć się jednym z algorytmów optymalizacji nieliniowej bez ograniczeń przedstawionych w dalszej części artykułu.

Warunki konieczne istnienia rozwiązania zadania optymalizacji z ograniczeniami postaci (3) noszą nazwę warunków Kuhna-Tuckera. Zanim zostaną one przedstawione, konieczne jest wprowadzenia pojęcia funkcji Lagrange'a.

Dla zadania (3) funkcja Lagrange'a ma następującą postać:

$$L(\mathbf{X}, \lambda) = f(\mathbf{X}) + \langle \lambda, \mathbf{g}(\mathbf{X}) \rangle \tag{4}$$

gdzie:
$\mathbf{g}(\mathbf{X}) = [g_1(\mathbf{X}), g_2(\mathbf{X}), \ldots, g_m(\mathbf{X})]$ jest wektorem ograniczeń,
$\lambda = [\lambda_1, \lambda_2, \ldots, \lambda_m]$ jest wektorem mnożników Lagrange'a.

Warunkiem koniecznym istnienia minimum lokalnego dla zadania programowania nieliniowego z ograniczeniami postaci (3) w punkcie $\mathbf{X}^*$ jest spełnienie następujących warunków:

1) funkcje f i g_i są różniczkowalne;
2) istnieje wektor $\lambda^* \geq \mathbf{0}$, taki że:

$$\begin{gathered} \nabla_x L(\mathbf{X}^*, \lambda^*) = \mathbf{0} \\ \langle \lambda^*, \nabla_\lambda L(\mathbf{X}^*, \lambda^*) \rangle = 0 \\ \nabla_\lambda L(\mathbf{X}^*, \lambda^*) \leq \mathbf{0} \end{gathered} \tag{5}$$

Warunek konieczny istnienia rozwiązania podany wyżej stanie się warunkiem wystarczającym, jeśli dodatkowo założona zostanie pseudowypukłość funkcji f i quasi-wypukłość wszystkich ograniczeń g_i. (Definicje funkcji pseudowypukłych i quasi-wypukłych podane są w literaturze [1], [2].)

Przedstawione wyżej warunki konieczne i wystarczające istnienia punktu optymalnego obowiązują dla zadania postaci (3), w którym wszystkie ograniczenia są nierównownościowe. Jeśli zadanie optymalizacji zawiera również ograniczenia równościowe, należy warunki te nieco zmodyfikować. Odpowiednie twierdzenia można znaleźć w literaturze [2].

W bardzo prostych przypadkach możliwe jest rozwiązanie zadania na bazie warunków K-T. W większości przypadków jest to niemożliwe i wówczas należy wykorzystać jeden z algorytmów optymalizacji nieliniowej z ograniczeniami.

Podklasą zadań optymalizacji z ograniczeniami jest zadanie programowania liniowego (lub optymalizacji liniowej). Dla takiego zadania funkcja celu i wszystkie ograniczenia muszą być funkcjami liniowymi. Zadanie optymalizacji liniowej rozwiązuje się metodą simpleks. Algorytm simpleks, podobnie jak metody optymalizacji nieliniowej, jest algorytmem iteracyjnym. Oznacza to, że w kolejnych krokach algorytmu otrzymuje się ciąg przybliżeń zbieżny do rozwiązania. Przewaga tego algorytmu nad algorytmami optymalizacji nieliniowej wynika ze sposobu wyznaczania kolejnego rozwiązania. U podstaw algorytmu simpleks leży następujące stwierdzenie: w przypadku liniowego zadania programowania matematycznego n-wymiarowa liniowa funkcja celu osiąga wartość minimalną lub maksymalną (jeśli ma jedno rozwiązanie) w jednym z wierzchołków n-wymiarowego wielościanu określonego przez ograniczenia. Aby znaleźć rozwiązanie, wystarczy więc sprawdzać wartość funkcji celu w tych punktach. W przypadku funkcji nieliniowej (bez względu na charakter ograniczeń) rozwiązanie optymalne może znajdować się w dowolnym miejscu, zarówno wewnątrz, jak i na granicy obszaru wyznaczonego ograniczeniami. Z tego też względu dla zadań nieliniowych należy spodziewać się większego nakładu obliczeń niż w przypadku wykorzystania algorytmu simpleks.

Algorytmy optymalizacji nieliniowej

Metody rozwiązywania zadań optymalizacji nieliniowej formułowane są w postaci algorytmów iteracyjnych.

Algorytm iteracyjny pozwala na wyznaczenie ciągu punktów $\mathbf{X}^0$, $\mathbf{X}^1$, $\mathbf{X}^2$, $\mathbf{X}^3$, ... , takich, że $\mathbf{X}^k \in R^n$ (lub $\mathbf{X}^k \in Z_R$, jeśli algorytm ma rozwiązywać zadanie optymalizacji z ograniczeniami). Każdy kolejny punkt $\mathbf{X}^k$ jest obliczany na

podstawie punktu poprzedniego (proces ten nazywany jest krokiem algorytmu). Algorytm musi być skonstruowany tak, aby generowany ciąg punktów $\{\mathbf{X}^k\}_{k=1,\ldots,\infty}$ coraz dokładniej przybliżał rozwiązanie zadania, lub inaczej mówiąc był zbieżny do tego rozwiązania. Algorytm jest zbieżny do rozwiązania określonego przez punkt $\mathbf{X}^*$, jeśli spełniony jest następujący warunek:

$$\lim_{k\to\infty} \mathbf{X}^k = \mathbf{X}^* \tag{6}$$

Warunek (6) zapewnia jedynie, że ciąg generowany przez algorytm jest zbieżny do rozwiązania w nieskończoności. Oczywiście nie oznacza to, że należy wykonać nieskończoną liczbę kroków. Dokładność obliczeń zakładana jest z góry i wówczas przybliżone rozwiązanie może być wyznaczone w skończonej liczbie kroków.

Schemat działania algorytmu iteracyjnego można przedstawić w następującej postaci:

1. Wyznacz punkt $\mathbf{X}^0$ (tzw. punkt startowy). Podstaw k = 0.
2. Oblicz $\mathbf{X}^{k+1}$ na podstawie $\mathbf{X}^k$.
3. Sprawdź, czy punkt $\mathbf{X}^{k+1}$ spełnia kryterium stopu. Jeśli tak, to koniec obliczeń (znaleziono rozwiązanie). Jeśli nie, to przejdź do kolejnego punktu.
4. Podstaw k = k + 1 i przejdź do punktu 2.

Kryterium stopu jest warunek, którego spełnienie zapewnia rozwiązanie zadania z wystarczającą dokładnością.

Metody rozwiązywania zadań optymalizacji nieliniowej bez ograniczeń

Metody optymalizacji bez ograniczeń dzielimy na metody poszukiwań prostych i metody poprawy.

W metodach poszukiwań prostych w kolejnych krokach algorytmu badane są wartości funkcji celu w otoczeniu aktualnego przybliżenia rozwiązania, czyli w otoczeniu punktu $\mathbf{X}^k$. Sposób wyboru badanych punktów z tego otoczenia zależy od zastosowanej metody. Jeśli wartość funkcji celu w którymś z tych punktów jest mniejsza niż w punkcie $\mathbf{X}^k$, dokonywana jest zmiana aktualnego przybliżenia rozwiązania. Jeśli nie znaleziono punktu, w którym następuje poprawa, zmniejszane jest przeszukiwane otoczenie, wobec czego poszukiwania odbywają się z większą dokładnością. Do omawianej grupy metod zaliczane są metody Rosenbrocka, Hooke'a-Jeevsa, Neldera-Meada [2]. Metody te są mało efektywne (znalezienie rozwiązania może wymagać wielu iteracji), ale bardzo

niezawodne. Ponieważ nie wymagają znajomości gradientu, mogą być wykorzystane w przypadkach, gdy funkcja celu jest nieróżniczkowalna.

W przypadku metod poprawy każdy krok algorytmu składa się z dwóch etapów. Pierwszy etap polega na wyznaczeniu kierunku poszukiwań, drugi jest minimalizacją funkcji celu wzdłuż tego kierunku. Znalezione minimum staje się kolejnym przybliżeniem rozwiązania.

Metody poprawy dzielone są na gradientowe i bezgradientowe. W przypadku metod gradientowych w kolejnych iteracjach wykorzystywana jest informacja o wartości funkcji celu i jej gradientu, natomiast w metodach bezgradientowych jedynie o wartości funkcji celu. Gradientowe metody poprawy nazywane są metodami kierunków poprawy. Metody gradientowe są zwykle znacznie efektywniejsze niż bezgradientowe. Dlatego też w przypadku zadań optymalizacji z różniczkowalną funkcją celu wskazane jest posługiwanie się metodami kierunków poprawy (czyli metodami gradientowymi). Jako przykłady bezgradientowych metod poprawy można podać metodę Powella czy Davisa-Swanna-Campeya [2], [3].

Wszystkie metody kierunków poprawy działają według następującego schematu:

1. Wyznacz punkt $\mathbf{X}^0$ (tzw. punkt startowy). Podstaw k = 0.
2. Wyznacz kierunek poprawy $\mathbf{D}^k$.
3. Znajdź punkt $\mathbf{X}^k$ będący minimum funkcji celu wzdłuż kierunku $\mathbf{D}^k$.
4. Sprawdź, czy punkt $\mathbf{X}^k$ spełnia kryterium stopu. Jeśli tak, to koniec obliczeń (znaleziono rozwiązanie). Jeśli nie, to przejdź do kolejnego punktu.
5. Podstaw k = k + 1 i przejdź do punktu 2.

Poszczególne metody różnić się będą głównie sposobem wyznaczania kierunku poprawy. W konkretnym algorytmie mogą być zastosowane różne metody minimalizacji w kierunku, jak również różne kryteria stopu.

Wszystkie metody poprawy wykorzystują algorytmy pozwalające znaleźć minimum funkcji celu w kierunku $\mathbf{D} \in R^n$. Zadanie to sprowadza się do wyznaczenia wartości parametru $\tau^* \geq 0$, gdzie $\tau^* \in R^1$, takiego, że:

$$f(\mathbf{X}_p + \tau^* \cdot \mathbf{D}) = \min_{\tau} f(\mathbf{X}_p + \tau \cdot \mathbf{D}) \tag{7}$$

gdzie: $\mathbf{X}_p \in R^n$ jest punktem początkowym minimalizacji w kierunku.

Znając τ^* można wyznaczyć punkt $\mathbf{X}_w \in R^n$ będący minimum funkcji f w kierunku $\mathbf{D}$ według zależności:

$$\mathbf{X}_w = \mathbf{X}_p + \tau^* \cdot \mathbf{D} \tag{8}$$

Metody minimalizacji w kierunku mogą być metodami bezgradientowymi (np. metoda złotego podziału, metoda interpolacji kwadratowej) i gradientowymi (metoda średnich geometrycznych, metoda interpolacyjno-ekstrapolacyjna). Opis wymienionych metod znaleźć można w literaturze [2], [3].

Kryterium stopu algorytmu optymalizacyjnego ma najczęściej postać:

$$\| \mathbf{X}^{k+1} - \mathbf{X}^{k} \| \leq \varepsilon \tag{9}$$

W nierówności (9) ε jest założoną dokładnością obliczeń, a $\| \mathbf{X}^{k+1} - \mathbf{X}^{k} \|$ to odległość przybliżonych rozwiązań uzyskiwanych w kolejnych iteracjach. Ponieważ spełnienie warunku stopu może wymagać bardzo dużego nakładu obliczeń, stosuje się dodatkowe zabezpieczenia. Przykładem może być założenie z góry maksymalnej liczby iteracji bądź obliczeń funkcji celu.

Najbardziej znane metody kierunków poprawy to metody najszybszego spadku, Newtona, gradientu sprzężonego i zmiennej metryki. Krótki opis wymienionych algorytmów zostanie przedstawiony poniżej.

W metodzie najszybszego spadku jako kierunek poszukiwań przyjmowany jest kierunek minus gradientu, czyli:

$$\mathbf{D}^{k} = -\nabla_x f(\mathbf{X}^{k}) \tag{10}$$

Następnie wykonywana jest minimalizacja wzdłuż tego kierunku, pozwalająca obliczyć punkt $\mathbf{X}^{k+1}$. W punkcie tym wyznaczany jest następnie gradient, który stanowi podstawę do określenia kolejnego kierunku, wzdłuż którego przeprowadzana zostanie minimalizacja.

Najczęściej stosowanym kryterium stopu jest warunek wyznaczony nierównością (9).

Istotną wadą omawianej metody jest wyraźne obniżenie szybkości zbieżności w wypadku funkcji, dla których minimum leży w tzw. wąskiej dolinie.

W metodzie Newtona kierunkiem poszukiwań jest kierunek minus gradientu pomnożony przez odwrotność hesjanu funkcji celu:

$$\mathbf{D}^{k} = -\mathbf{H}^{-1}(\mathbf{X}^{k}) \cdot \nabla_x f(\mathbf{X}^{k}) \tag{11}$$

gdzie: $\mathbf{H}(\mathbf{X}^{k})$ jest wartością hesjanu w punkcie $\mathbf{X}^{k}$.

Metoda Newtona charakteryzuje się szybką zbieżnością do rozwiązania. Istotną jej wadą jest konieczność wyznaczania i odwracania w każdym kroku hesjanu. Jak wiadomo, w przypadku dużego wymiaru macierzy poszukiwanie jej odwrotności sprawia istotne trudności numeryczne.

W literaturze spotkać można wiele odmian metody gradientów sprzężonych. Wspólną ich cechą jest to, że kierunki poszukiwań generowane w kolejnych krokach są kierunkami sprzężonymi. Definicję kierunków sprzężonych można znaleźć w literaturze (2).

W pierwszym kroku algorytmu jako kierunek poszukiwań minimum przyjmowany jest kierunek minus gradientu. Minimum funkcji staje się przybliżeniem rozwiązania w kroku następnym. W k-tym kroku kierunek poszukiwań określany jest według reguły:

$$\mathbf{D}^k = -\nabla_x f(\mathbf{X}^k) + \beta_k \cdot \mathbf{D}^{k-1} \tag{12}$$

gdzie β_k jest współczynnikiem.

Poszczególne odmiany metod gradientu sprzężonego różnią się sposobem wyznaczania współczynnika β. Odpowiednie wzory podane są w literaturze (2).

W metodzie, a właściwie grupie metod zmiennej metryki w kolejnych krokach algorytmu generowany jest ciąg macierzy $\{\mathbf{V}^k\}$ będących przybliżeniami odwrotności hesjanu. Ciąg ten wyznaczany jest na podstawie zmian gradientu funkcji celu w poprzednio wykonanych krokach. Metody zmiennej metryki są podobne do metody Newtona i dlatego nazywane są metodami quasi-newtonowskimi.

W każdym kroku algorytmu tworzony jest kierunek poszukiwań według zależności:

$$\mathbf{D}^k = -\mathbf{V}_k \cdot \nabla_x f(\mathbf{X}^k) \tag{13}$$

Następnie w wyniku minimalizacji funkcji w tym kierunku otrzymywany jest punkt $\mathbf{X}^{k+1}$.

Poszczególne odmiany metody zmiennej metryki różnią się sposobem wyznaczania macierzy $\mathbf{V}_k$. Wzory te są dość złożone i z tego względu nie zostaną tu przedstawione – można znaleźć je w literaturze (2).

Metody rozwiązywania zadań optymalizacji nieliniowej z ograniczeniami

Metody optymalizacji nieliniowej z ograniczeniami zaliczane są do metod aproksymacyjnych. Zamiast zadania wyjściowego określonego zależnością (3) rozwiązywany jest ciąg zadań zastępczych. Ciąg ten tworzony jest w taki sposób, aby zadania zastępcze w kolejnych iteracjach coraz lepiej przybliżały zadanie wyjściowe.

Najbardziej znanymi grupami metod optymalizacji nieliniowej z ograniczeniami są metody funkcji kary, metody modyfikacji kierunku i metoda aproksymacji kwadratowej.

W metodach funkcji kary dokonywana jest zamiana zadania optymalizacji z ograniczeniami na zadanie bez ograniczeń. Polega ona na zmodyfikowaniu funkcji celu przez wprowadzenie do niej kary za przekroczenie ograniczeń. Poszczególne odmiany metod różnią się sposobem wprowadzenia kary.

W metodzie przesuwanej funkcji kary opracowanej przez Powella funkcja celu modyfikowana jest w następujący sposób:

$$P(\mathbf{X}, \mathbf{A}, \mathbf{U}) = f(\mathbf{X}) + \frac{1}{2} \cdot \sum_{i=1}^{m} \alpha_i \cdot [\max(0; g_i(\mathbf{X}) + u_i)]^2 \qquad (14)$$

Zadanie zastępcze polegać będzie na minimalizacji funkcji $P(\mathbf{X}, \mathbf{A}, \mathbf{U})$ względem $\mathbf{X}$ zamiast funkcji $f(\mathbf{X})$. Wektor $\mathbf{A} = [\alpha_1, \alpha_2, \ldots, \alpha_m]$ nazywany jest wektorem współczynników funkcji kary. Współczynniki te muszą być dodatnie. Wektor $\mathbf{U} = [u_1, u_2, \ldots, u_m]$ natomiast jest wektorem przesunięć kary. Wymiar wektorów $\mathbf{A}$ i $\mathbf{U}$ równy jest liczbie ograniczeń. Wektory te mają ustalone wartości w każdym kroku, po czym są tak modyfikowane, aby ciąg $\{\mathbf{X}^k\}$ dążył przy $k \to \infty$ do rozwiązania zadania (3).

Wprowadzenie kary za przekroczenie ograniczeń do funkcji celu pozwala zastąpić zadanie optymalizacji z ograniczeniami zadaniem bez ograniczeń. Zmodyfikowane zadanie jest następnie rozwiązywane metodami optymalizacji nieliniowej bez ograniczeń.

Metody z modyfikacją kierunku dzielą się na metody kierunków dopuszczalnych i metody rzutu ortogonalnego ([2], [3]). Różnią się one sposobem wyznaczania kierunku, w którym minimalizowana jest funkcja celu. W pierwszym przypadku w otoczeniu ograniczeń generowane są kierunki dopuszczalne (przynajmniej częściowo mieszczące się w obszarze rozwiązań dopuszczalnych), w drugim natomiast kierunek gradientu rzutowany jest na powierzchnię styczną do ograniczeń.

W metodzie aproksymacji kwadratowej zadanie optymalizacji nieliniowej z ograniczeniami zastępowane jest w poszczególnych krokach odpowiednio dobranym zadaniem programowania kwadratowego (kwadratowa funkcja celu i liniowe graniczenia). Reguły tworzenia zadania zastępczego znaleźć można w literaturze ([2], [3]).

Problemy związane ze stosowaniem algorytmów optymalizacji nieliniowej

Zastosowanie konkretnego algorytmu optymalizacji nieliniowej wymaga od użytkownika podania wartości parametrów sterujących oraz punktu startowego. Do grupy parametrów sterujących zaliczamy takie parametry, jak dokładność obliczeń czy maksymalna liczba wyznaczeń wartości funkcji celu. W metodach kierunków poprawy należy dodatkowo podać dokładność wyznaczania minimum w kierunku.

Parametry związane z dokładnością w istotny sposób wpływają na przebieg procesów obliczeniowych. Zbyt duża ich wartość powoduje wydłużenie czasu obliczeń, jak również z powodu błędów numerycznych może w pewnych wypadkach uniemożliwić rozwiązanie. Za mała natomiast może prowadzić do błędnego rozwiązania. Dokładność powinna być odpowiednio dobrana do stosowanej metody i do konkretnego zadania. Zwykle doboru dokonuje się metodą prób i błędów. W pewnych przypadkach może wystąpić konieczność uruchamiania algorytmu kilkakrotnie z różnymi wartościami parametrów. Wówczas obowiązuje następująca zasada: im bliżej rozwiązania wykonywane są obliczenia, tym większa powinna być dokładność. Wyznaczone punkty – najlepiej przybliżające rozwiązanie – stają się punktami startowymi w kolejnych uruchomieniach algorytmu.

Dobór punktu startowego ma również bardzo duże znaczenie. Im punkt ten znajduje się bliżej rozwiązania, tym większa gwarancja pomyślnego zakończenia obliczeń. Wiele metod wymaga poza podaniem punktu startowego oszacowania odległości od rozwiązania. Należy podać wielkość promienia kuli o środku w punkcie startowym, w której powinno znajdować się rozwiązanie. Jest to szczególnie trudne, ponieważ zwykle nie znamy rozwiązania zadania nawet w sposób przybliżony. Promień należy wówczas określić przeprowadzając eksperymenty polegające na wielokrotnym uruchamianiu algorytmu z różnymi jego wartościami. Obserwacja wyników obliczeń pozwoli ocenić, jaka wartość jest najbardziej odpowiednia. W przypadku zadań z ograniczeniami należy ponadto zadbać, aby punkt startowy znajdował się w obszarze rozwiązań dopuszczalnych.

Kolejna grupa problemów może wynikać ze złego uwarunkowania zadania optymalizacji. Zadanie optymalizacji bez ograniczeń jest źle uwarunkowane, jeśli różnica między najmniejszą a największą wartością hesjanu funkcji celu w pobliżu rozwiązania jest duża. Konsekwencją tego są bardzo duże różnice w przyrostach funkcji celu spowodowanych jednakowymi przyrostami zmien-

nych decyzyjnych. Przyrost wartości funkcji w wyniku zmiany wartości jednej zmiennej może być o kilka rzędów wielkości większy niż spowodowany taką samą zmianą innej zmiennej. W wypadku złego uwarunkowania należy przeprowadzić skalowanie zmiennych. Skalowanie polega na takim doborze jednostek, w jakich wyrażane są zmienne decyzyjne, aby ich wpływ na wartość funkcji celu był podobny.

W przypadku zadań optymalizacji z ograniczeniami o złym uwarunkowaniu zadania mówi hesjan funkcji Lagrange'a. Złe uwarunkowanie może w tym przypadku przejawiać się dużymi różnicami we wpływie poszczególnych zmiennych decyzyjnych na wartości funkcji celu, jak również na ograniczenia. W takich wypadkach należy poza skalowaniem zmiennych zastosować skalowanie ograniczeń, które polega na pomnożeniu ograniczeń przez właściwie dobrane wielkości.

Literatura

1. CHIANG A., 1994: *Podstawy ekonomii matematycznej*. PWE, Warszawa.
2. FINDEISEN W., SZYMANOWSKI J., WIERZBICKI A., 1977: *Teoria i metody obliczeniowe optymalizacji*. PWN, Warszawa.
3. KRĘGLEWSKI T., ROGOWSKI T., RUSZCZYŃSKI A., SZYMANOWSKI J., 1984: *Metody optymalizacji w języku FORTRAN*. PWN, Warszawa.
4. PANEK E., 1993: *Elementy ekonomii matematycznej. Statyka*. Wydawnictwo Naukowe PWN, Warszawa.

The elements of optimisation theory and methods of solving non-linear optimisation problems

Abstract

This article contains the elements of optimisation theory – the mathematical forms of nonlinear optimisation problems without constraints and with constraints and preconditions and sufficient conditions for existence of the optimal solution.

Next the author presents methods of non-linear optimisation. This presentation includes methods of solving problems without constraints and with constraints. Difficulties occurred when using these methods are discussed at the end.

Andrzej Kluza
Katedra Ekonometrii i Informatyki, SGGW

Model sieciowy wykorzystania maszyn rolniczych przy wykonywaniu prac agrotechnicznych

Wstęp

Opracowanie przedstawia próbę modelowania sytuacji wykorzystania maszyn rolniczych w celu wyznaczenia optymalnego składu parku maszynowego gospodarstwa rolniczego za pomocą rozwiązania sieciowego zadania optymalizacji. Kryterium optymalizacji stanowi najmniejszy całkowity koszt wykonania zabiegów agrotechnicznych. Optymalizacji indywidualnego [6, 7] i zespołowego [2, 4] wykorzystania parku maszynowego dokonuje się najczęściej korzystając z różnorodnych odmian programowania liniowego [2]. Zastosowanie do tego celu programowania sieciowego ze względu na niektóre jego korzystne cechy, jak duża szybkość obliczeń i prostota modelu [1], może budzić nadzieje na stworzenie bardziej efektywnego narzędzia optymalizacji.

W pracy wykorzystano dane empiryczne zebrane w ramach tematu „System gromadzenia i przetwarzania danych na potrzeby doskonalenia i zarządzania gospodarstwami rolnymi" realizowanego przez Katedrę Ekonomiki i Organizacji Przedsiębiorstw Rolniczych SGGW [2].

Model relacji między obiektami

Elementami składowymi wykonywania prac agrotechnicznych są maszyny rolnicze, mające pewne ograniczone zasoby czasu pracy i określony koszt jednostkowy wykonania pracy, techniki wykonania zbiegów agrotechnicznych oraz obszary pól uprawnych, na których wykonuje się zabiegi.

Dane empiryczne zostały przeanalizowane i doprowadzono je do czwartej postaci normalnej relacyjnej struktury danych. Podstawą modelu sieciowego są wypływające stąd logiczne relacje jakościowe pomiędzy obiektami. Ustalenie zależności ilościowych jest uzależnione od konkretnego typu ścieżki wybranego do modelowania.

Dla uproszczenia w podanym poniżej modelu uwzględnia się jeden okres agrotechniczny, obejmujący od kilku do kilkudziesięciu dni zależnie od natężenia prac polowych.

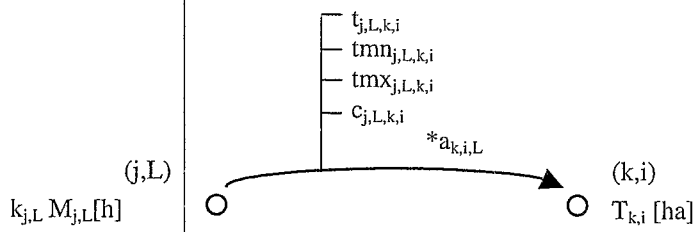
Maszyny rolnicze wykonujące prace polowe podzielono na grupy maszyn mogących wykonywać tę samą operację. Do grup maszyn należą trzy typy obiektów uniwersalnych: ciągniki, przyczepy i zasoby pracy własnej pracowników, oraz poszczególne typy maszyn roboczych wykonujących szczegółowe prace uprawowe, np. siewniki, sadzarki. Zabiegi agrotechniczne wykonywane są za pomocą poszczególnych technik uprawowych. Do wykonania jednej techniki trzeba pracy wielu typów maszyn jednocześnie. Dany zabieg agrotechniczny może być wykonany za pomocą różnych technik. Zapotrzebowanie na wykonanie zabiegu wyrażone w [ha] zaspokajane jest dzięki sumarycznemu wykonaniu prac polowych różnymi technikami.

Model łuku pracy maszyny rolniczej

W przypadku wykonywania prac agrotechnicznych można użyć analogii przepływu. Przepływać przez sieć może dostępny czas pracy od maszyn rolniczych do technik wykonania prac polowych, które z kolei zaspokajają zapotrzebowanie na wykonanie zabiegów agrotechnicznych ustalane dla poszczególnych pól zgodnie z planami agrotechnicznymi. Graf, na którego łukach określona jest wartość funkcji, nazywamy siecią. Funkcją opisana na łukach jest koszt jednostkowego czasu pracy maszyny. Przy uwzględnieniu powyższych założeń problem wykorzystania czasu pracy maszyn rolniczych można wyrazić jako problem najtańszego przepływu w sieci.

W celu przekształcenia przepływu z jednostek dostępnego czasu pracy maszyn rolniczych na jednostki powierzchni pól, na których wykonano pracę polową daną techniką, należy zastosować mnożnik na łuku przepływu pomiędzy maszyną a techniką dający możliwość konwersji jednostek przepływu czasu pracy maszyny [h] na wartość obszaru wykonanej pracy [ha], mający znaczenie wydajności pracy maszyny (rys. 1).

Łuk wykonania pracy przez daną maszynę opisany jest parą: ((j,L),(k,i)), gdzie ciąg (j,L) oznacza węzeł początkowy, zaś ciąg (k,i) – węzeł końcowy w zbiorze połączeń transportowych U grafu G. Każda maszyna jako dostawca czasu pracy $k_{j,L}M_{j,L}$ [h] jest przedstawiona jako węzeł i jest połączona z każdym węzłem techniki $T_{k,i}$ [ha], przedstawiającym wykonanie pracy w k-tej technice i-tego zabiegu.



Rysunek 1
Opis łuku sieci przepływowej pomiędzy węzłem przedstawiającymi zasób czasu pracy maszyny a węzłem przedstawiającym wykonanie pracy polowej na danym obszarze za pomocą danej techniki
Źródło: Opracowanie własne.

Wprowadźmy oznaczenia:

j – numer maszyny w grupie maszyn, j = 1, ..., R_L,
R_L – liczba maszyn w L-tej grupie maszyn,
L = 1, ..., S – numer grupy maszyn,
$k_{j,L}$ – stosunek wydajności j-tej maszyny z L-tej grupy maszyn do j = 1 pierwszej maszyny w tej grupie, mającej najmniejszą wydajność,
$M_{j,L}$ – ilość dostępnych zasobów czasu pracy [h] j-tej maszyny z L-tej grupy maszyn.

Wartości zasobów pracy poszczególnych maszyn jest przypisana do węzłów początkowych sieci.

Obszar pracy wykonanej przez maszynę w określonej technice przedstawiony jest iloczynem:

$$T_{k,i}\ [ha] = k_{j,L} M_{j,L}\ a_{k,i,L}\ [h][ha/h],$$

gdzie:
i – numer zabiegu agrotechnicznego, i = 1, ..., V,
V – liczba wszystkich rodzajów zabiegów agrotechnicznych,
k = 1,..., P_V – numer techniki agrotechnicznej,
$a_{k,i,L}$ – zapotrzebowanie k-tej techniki na czas pracy pierwszej maszyny z L-tej grupy maszyn przy wykonywaniu i-tego zabiegu agrotechnicznego [ha/h].

Zapotrzebowanie Z_i [ha] na określony zabieg agrotechniczny jest sumą zapotrzebowań na ten zabieg z poszczególnych pól.

Na łuku opisane są cztery wartości. Trzy z nich są dane:

$tmn_{j,L,k,i}$ – czas pracy maszyny [h] ustalony jako minimalny,
$tmx_{j,L,k,i}$ – czas pracy maszyny [h] ustalony jako maksymalny,
$c_{j,L,k,i}$ – koszt jednostkowy czasu pracy maszyny [zł/h].

W wyniku działania algorytmu sieciowego wyznaczana jest czwarta wartość, określona na łukach sieci między węzłem początkowym a końcowym:
$t_{j,L,k,i}$ – czas pracy maszyny [h] pracującej w k-tej technice.

Dyskusja proporcjonalności przepływów pracy

Ogólny model ścieżki wykonania pracy maszyną rolniczą w danej technice, jaki przedstawiono na rysunku 1, nie uwzględnia całej złożoności operacji na danych empirycznych, wynikającej z rzeczywistego sposobu użytkowania maszyn rolniczych w czasie prowadzenia prac uprawowych.

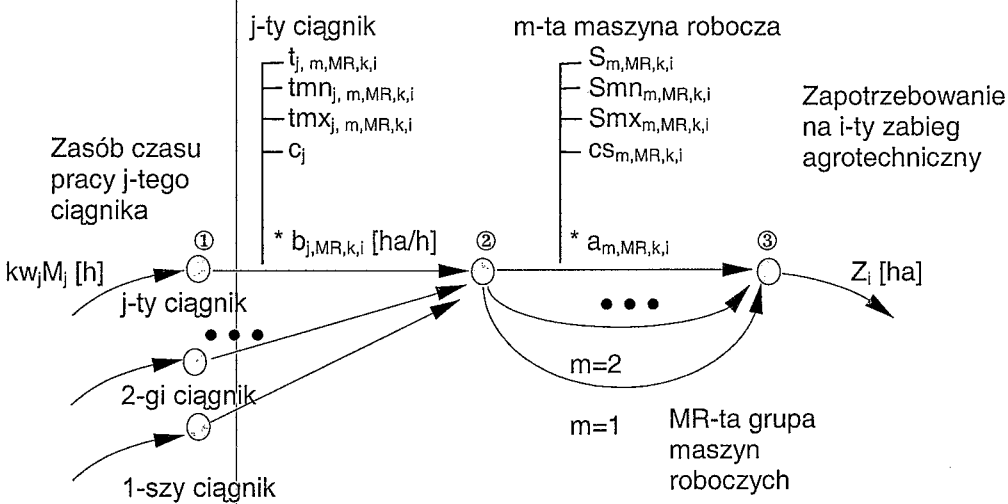
Na każdą z technik agrotechnicznych składa się praca różnych maszyn stanowiących zespół roboczy, składający się zwykle z ciągnika, maszyny roboczej i obsługi traktowanej jako „maszyna". Każda z maszyn zużywa proporcjonalne ilości czasu pracy w stosunku do obszaru wykonywanej pracy. Wynika stąd wniosek, iż czas pracy maszyn z zespołu zużywany jest w niezmiennych proporcjach. W niektórych przypadkach czas pracy jednej maszyny, konieczny do wykonania pracy na 1 ha pola, różni się od takiego czasu pracy innej maszyny z zespołu. Występuje tu więc konieczność uwzględnienia proporcjonalności pomiędzy zużytkowanymi czasami pracy poszczególnych maszyn przy modelowaniu przepływów czasu pracy.

Gdyby do optymalizacji pracy maszyn użyć sieci mającej początki łuków wykonania pracy w węzłach zasobów pracy poszczególnych maszyn wchodzących do zespołu pracującego w danej technice, a końce łuków w węzłach zapotrzebowania na zabiegi, okazałoby się, że proporcjonalne wykorzystanie maszyn nie nastąpiłoby. Działoby się tak dlatego, iż algorytm optymalizujący według najtańszego przepływu nie „widziałby" różnicy w korzystaniu z czasu pracy poszczególnych maszyn i wyznaczyłby zużycie czasu pracy maszyny o najmniejszym koszcie. Mogłoby się więc zdarzyć, że w pracy polowej miałaby uczestniczyć tylko jedna maszyna robocza, bez obsługi i bez ciągnika. Aby

temu zaradzić, należy stworzyć taki model wykorzystania czasu pracy, który uwzględniałby proporcjonalne wykorzystanie czasu pracy maszyn wchodzących do zestawu.

Ścieżka wykonania pracy polowej

Pozioma ścieżka sieci na rysunku 2 reprezentuje wykonywanie i-tego zabiegu agrotechnicznego w k-tej technice z proporcjonalnym wykorzystaniem czasu pracy zespołu maszyn. Łuk ① ② ścieżki reprezentuje pracę ciągnika w zespole, a łuk ② ③ pracę zespołu: maszyna robocza, praca własna ludzi, ewentualnie przyczepa, oraz ewentualnie druga maszyna robocza.



Rysunek 2
Ścieżki sieci proporcjonalnego wykorzystania czasu pracy maszyn roboczych i ciągników wykonujących pracę k-tą techniką w i-tym zabiegu w ramach pracy w zespole maszyn
Źródło: Opracowanie własne.

W przypadku, gdy w MR-tej grupie maszyn roboczych znajduje się do dyspozycji więcej niż jedna maszyna oraz gdy do wykonywania prac w danej technice zdatny i dostępny jest więcej niż jeden ciągnik, dodajemy pozostałe ścieżki przepływu. Sumowanie pracy wszystkich ciągników odbywa się

w węźle ① w jednostkach obszaru [ha] wypracowanej przez ciągnik pracy polowej. Dzięki sumowaniu obszaru wykonywanej pracy uwzględnione są różne wydajności i koszty eksploatacji ciągników różnych typów mogących uczestniczyć w pracach w tej technice. Mnożnik bj,MR,k,i na łuku pomiędzy węzłami ① i ② stwarza proporcjonalność przepływu pomiędzy czasem pracy ciągnika a czasem pracy reszty maszyn z zespołu, zamieniając ilość czasu pracy ciągnika na liczbę ha powierzchni wykonanej pracy daną techniką przez ciągnik. Innymi słowy, aby do węzła ② dopłynął 1 ha możliwej do wykonania pracy w danej technice, z węzła ① musi wypłynąć tyle godzin [h] czasu pracy ciągnika, ile wynika z zapotrzebowania maszyny roboczej i przyczepy na pracę ciągnika w tej technice. Przy zapotrzebowaniu 10 h/ha na ciągnik ilość możliwej do wykonania pracy wynosi 0,1 ha na 1 h pracy ciągnika.

Łuki pomiędzy węzłami ② ③ reprezentują różne m-te maszyny z MR-tej grupy maszyn roboczych pracujące w zespołach roboczych w k-tej technice w celu wykonania i-tego zabiegu. Łączny obszar wykonania zabiegu poszczególnymi technikami jest sumowany w węźle ③.

Oznaczenia:

M_j – dostępny czas pracy j-tego ciągnika [h],

$t_{j,m,MR,k,i}$ – ilość czasu pracy ciągnika w k-tej technice wykonania i-tego zabiegu [h],

kw_j – współczynnik wydajności j-tego ciągnika w stosunku do pierwszego ciągnika w grupie,

c_j – koszt jednostkowym eksploatacji j-tego ciągnika [zł/h],

MR = 4,.., L – numer grupy maszyn roboczych, które nie są ciągnikami, przyczepami ani zasobami pracy własnej,

L – liczba wszystkich grup maszyn.

Rozważmy pierwszy łuk ① ②. Aby ustalić proporcjonalność wykorzystania czasu pracy, przyjmijmy oznaczenia:

$Z_{Tm,MR,k,i}$ – jednostkowe zapotrzebowanie na czas pracy m-tej maszyny roboczej z MR-tej grupy maszyn przy wykonywaniu k-tą techniką i-tego zabiegu [h/ha],

$ZT_{2,k,i}$ – jednostkowe zapotrzebowanie na czas pracy przyczepy przy wykonywaniu k-tą techniką i-tego zabiegu [h/ha].

$ZC_{m,MR,k,i}$ – jednostkowe zapotrzebowanie m-tej maszyny roboczej z MR-tej grupy maszyn na czas pracy ciągnika przy wykonywaniu k-tą techniką i-tego zabiegu [h/h].

$ZC_{2,k,i}$ – jednostkowe zapotrzebowanie przyczepy na czas pracy ciągnika przy wykonywaniu k-tą techniką i-tego zabiegu [h/h].

Zapotrzebowanie maszyny roboczej na ciągnik wyrażone w godzinach pracy ciągnika na ha pracy polowej wynosi $ZT_{m,MR}\ ZC_{m,MR}$, dla przyczepy zaś $ZT_2\ ZC_2$. W sumie w tej technice wykonywania pracy polowej na 1 ha wykonania pracy polowej potrzeba $ZT_{m,MR}\ ZC_{m,MR} + ZT_2\ ZC_2$ godzin pracy ciągnika. Zatem ciągnik może wykonać $b_{j,MR,k,i} = \dfrac{1}{ZT_{m,MR,k,i} ZC_{m,MR,k,i} + ZT_{2,k,i} ZC_{2,k,i}}$ [ha] pracy polowej w czasie 1 godziny swej pracy.

Całkowity czas pracy ciągnika $t_{j,m,MR}$ jest mnożony przez współczynnik $b_{j,MR,k,i}$ [ha/h], co daje wykonanie $S_{m,MR,k,i}$ hektarów wykonania pracy polowej w tej technice.

Łuk ② ③ ma znaczenie wykonywania pracy przez zespół maszyn bez uwzględnienia ciągnika. Łuk modelujący wykonywanie przez m-tą maszynę z MR-tej grupy maszyn k-tej techniki w i-tym zabiegu agrotechnicznym wychodzi z węzła maszyny roboczej ② i kończy się w węźle ③ reprezentującym i-ty zabieg Z_i. Maksymalny posiadany zasób czasu pracy tej maszyny $tmx_{m,MR,k,i}$ jest ustalany przez ograniczenie górne przepływu łuku ② ③:

$$Smx_{m,MR,k,i} = \frac{tmx_{m,MR,k,i}}{ZT_{m,MR,k,i}} \text{ [ha]}$$

Koszt jednostkowy $cs_{m,MR,k,i}$, jakim jest obciążony przepływ obszaru pracy maszyn w zespole z wyłączeniem ciągnika $S_{m,MR,k,i}$, jest sumą kosztów jednostkowych zużycia czasu pracy własnej, kosztu eksploatacji maszyny roboczej i kosztu eksploatacji przyczepy, o ile jest ona wykorzystywana w tej technice. Koszt ten jest obliczany w złotych na ha wykonanej pracy:

$$cs_{m,MR,k,i} = c_{1pw}\ ZTpw_{,k,j} + c_{m,MR}\, ZT_{m,MR} + c_2\, ZT_{2,}$$

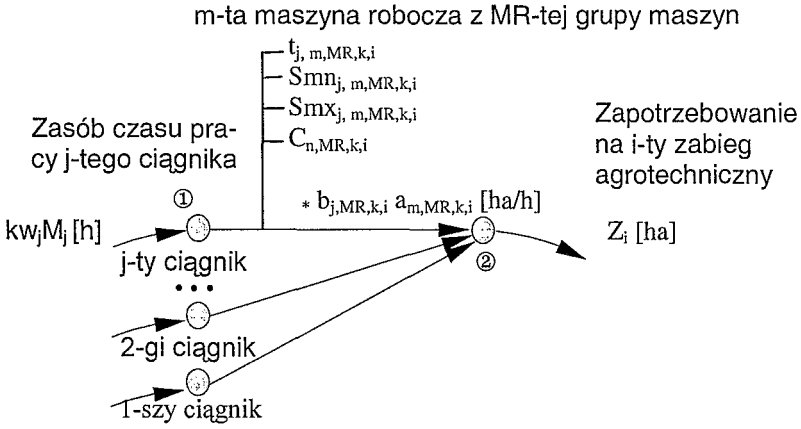
gdzie:

$ZTpw_{MR}$ – jednostkowe zapotrzebowanie na czas pracy ludzkiej przy wykonywaniu k-tą techniką i-tego zabiegu [h/ha].
c_{1pw} – jednostkowy koszt czasu pracy własnej pracowników rolnych [zł/h],
$c_{m,MR}$ – jednostkowy koszt eksploatacji maszyny roboczej [zł/h],
c_2 – jednostkowy koszt eksploatacji przyczepy [zł/h].

Po węźle reprezentującym zasoby czasu pracy maszyny roboczej na łuku dochodzącym do węzła zabiegu agrotechnicznego występuje mnożnik przepływu na łuku $a_{m,MR,k,i}$. Maszyna robocza posiadająca większą wydajność niż pierwsza maszyna z MR-tej grupy maszyn ma możliwość wykonać $kw_{m,MR}$ więcej pracy. Stąd mnożnik przepływu obszaru pracy:

$$a_{m,MR,k,i} = kw_{m,MR}$$

Większość maszyn ma zastosowania w wielu technikach. W tym przypadku można utworzyć ścieżki pracy maszyn bez możliwości bilansowania czasu ich pracy przez wykonujący optymalizację algorytm (rys. 3).



Rysunek 3
Ścieżki proporcjonalnego wykorzystania czasu pracy maszyny roboczej z niebilansowanym czasem pracy i ciągnika w zespole roboczym w k-tej technice i-tego zabiegu agrotechnicznego
Źródło: Opracowanie własne.

Pracę maszyny niebilansowanej wraz z pracą własną i ewentualnie przyczepą symbolizuje łuk ① ②. Jedynymi dostępnym ograniczeniami są dolne i górne ograniczenie obszaru wykonanych prac w technice: $Smn_{j,m,MR,k,i}$, $Smx_{j,m,MR,k,i}$. Bilans całego zaplanowanego przez algorytm czasu pracy maszyny można porównać z dostępnym czasem pracy maszyn tego typu już po dokonaniu obliczeń. Jednak jest niemożliwe nadanie warunku, aby algorytm czerpał z zapasów czasu pracy takiej maszyny tylko do określonej wartości. Uwzględ-

niając technikę z niebilansowaną maszyną, można otrzymać rozwiązania uwzględniające wykonywanie prac bez ograniczeń lub niewykonywanie ich wcale.

$$C_{n,MR,k,i} = \frac{c_j (ZT_{m,MR,k,i} ZC_{m,MR,k,i} + ZT_{2,k,i} ZC_{2,k,i}) + c_{m,MR} ZT_{m,MR,k,i} + c_2 ZT_{2,k,i} + c_{pw} ZT_{pw,k,i}}{ZT_{m,MR,k,i} + ZT_{2,k,i}}$$

Oznaczenia:

$C_{n,MR,k,i}$ – koszt jednostkowy wykonywania pracy przez zespół z n-tą niebilansowaną maszyną z MR-tej grupy przy pracy w k-tej technice i-tego zabiegu [zł/h],

gdzie:

$ZC_{m,MR,k,i}$ – zapotrzebowanie m-tej maszyny z MR-tej grupy na pracę ciągnika przy pracy w k-tej technice i-tego zabiegu [h/h],

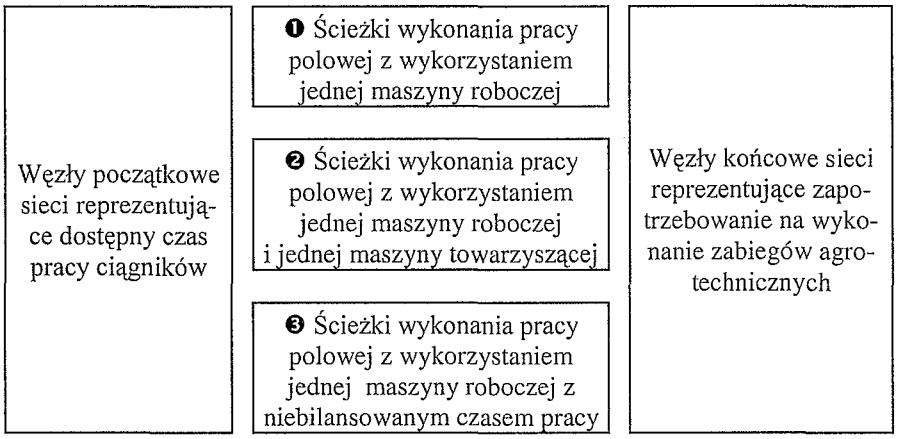
$ZC_{2,k,i}$ – zapotrzebowanie przyczepy na pracę ciągnika przy pracy w k-tej technice i-tego zabiegu [h/h].

Zbiorczy model sieciowy

Zbiorczy model sieciowy wykonywania zabiegów agrotechnicznych (rys. 4) składa się ze wszystkich typów ścieżek, rozpoczynających się w węzłach dostępnego czasu pracy wszystkich posiadanych ciągników i kończących się w węzłach sumujących wykonane prace polowe w obrębie danego zabiegu agrotechnicznego. Maszyny robocze mogące mieć zbilansowany czas pracy są modelowane ścieżkami typu ❶, Maszyny o bilansowanym czasie pracy pracujące wraz z maszynami niebilansowanymi są modelowane ścieżkami typu ❷. Maszyny robocze pracujące w wielu technikach w ciągu danego okresu agrotechnicznego modelowane są za pomocą ścieżek typu ❸. Poszczególne techniki wykonywania jednego zabiegu agrotechnicznego mogą być modelowane różnymi typami ścieżek przepływu, w zależności od tego, jakie maszyny będą użyte i w jakiej konfiguracji.

Otrzymany model został zweryfikowany i rozwiązany za pomocą oprogramowania optymalizującego sieci uogólnione [3] na podstawie danych empirycznych. Warianty obliczeń obejmowały ustalenie możliwości obniżenia kosztów mechanizacji przy zespołowym użytkowaniu maszyn. Wyliczone obniżenie kosztów wahało się od 12 do 25% całkowitych kosztów prac mechanizacyjnych. Ta wartość mieści się w przedziale do maksimum 30%, będącym

typowym rezultatem otrzymywanym z użyciem optymalizacji z zastosowaniem programowania liniowego [2], badań analitycznych [4], oraz jako wartość rzeczywistych rezultatów w rolnictwie niemieckim [4]. Pozytywny wynik obliczeń może potwierdzać użyteczność zaprezentowanego modelu.



Rysunek 4
Zbiorczy model sieci wykonywania zabiegów agrotechnicznych
Źródło: Opracowanie własne.

Model sieciowy wykorzystania maszyn oparty jest na prostej strukturze, lecz jego konstrukcja optymalizacji jest skomplikowana, a obsługa programu optymalizacyjnego jako aplikacji laboratoryjnej wymaga dużego nakładu pracy. Sam algorytm optymalizacji sieciowej odznacza się zaś dużą prędkością obliczeniową w odróżnieniu, od algorytmów rozwiązujących zadania programowania liniowego.

Literatura

1. AHUJA R.K., MAGNANTI T.L., ORLIN J.B., 1993: *Network Flows. Theory, Algorithms, and Applications*, Prentice Hall, Englewood Cliffs, New Jersey.
2. BORKOWSKI B., 1994: Metoda racjonalnego wykorzystania maszyn rolniczych w gospodarstwach chłopskich z zastosowaniem pakietu programowego MSERVICE/4. *Rozprawy Naukowe i Monografie*, Wydawnictwo SGGW, Warszawa.

3. CURET N., 1998: Implementation of a Steepest-Edge Primal-Dual Simplex Method for Network Linear Programs. *Annals of Operations Research* 81: 251–270.
4. KADNER K., 1995: *Zespołowe użytkowanie maszyn rolniczych. Doradztwo specjalistyczne. Podręcznik doradcy*. Messel.
5. MUZALEWSKI A., WÓJCICKI Z., 2000: *Ekonomiczno-organizacyjne aspekty zespołowego użytkowania maszyn rolniczych*. Warszawa, IBMER.
6. PAWLAK J., 1997: *Dobór maszyn i ich racjonalne użytkowanie*. IBMER, Warszawa.
7. WÓJCICKI Z., 1996: *Wyposażenie rolnictwa w środki techniczne – stan i kierunki przemian w układzie sektorowym i regionalnym*. IBMER, Warszawa.

Generalized network model of agriculture machinery operation

Abstract

Some of ways of reducing farm machines operating costs are rational machinery assemblage composition and optimal farm jobs assignment. Optimization solution for this tasks can be done by solving network programing task. Relations between machines and field work lead to a relational database data model. A basic generalized network path of many machines working simultaneously is created, and all quantitative dependencies are established. Depending on field work type three types of path are identified, and incorporated into final cumulative network model. This structure has served the autor for optimization calculations on real empirical data, which can attest usfulness of the model.

Stanisław Jabłonowski
Katedra Ekonometrii i Informatyki SGGW

Kontrakty terminowe w strategiach zabezpieczania się przed niekorzystnymi zmianami cen na Warszawskiej Giełdzie Towarowej

Wstęp

Warszawska Giełda Towarowa, działająca od 1995 roku, ma na celu promowanie i organizowanie nowoczesnych technik obrotu towarami, instrumentami finansowymi i prawami do tychże. Duży nacisk położono na rozwój transakcji terminowych i opcyjnych, które są jednymi z ważniejszych instrumentów pochodnych, zwanych także derywatami. W zamierzeniach początkowo miały rozwijać się przede wszystkim obroty towarami rolno-spożywczymi. W chwili obecnej spośród kontraktów terminowych i opcyjnych dobrze rozwijają się jedynie te pierwsze, i to nie na produkty rolne (choć można je zawierać na pszenicę i żywiec wieprzowy), ale przede wszystkim kontrakty terminowe na dolara amerykańskiego (dokładniej: na kurs USD; można jeszcze zawierać kontrakty na kurs EUR i na stopy procentowe). Tutaj występuje największa płynność (ponad sto kontraktów na kurs USD dziennie), która ma zasadnicze znaczenie w interesującym nas przypadku zabezpieczania się przed stratami finansowymi. W niniejszym artykule omawiam schemat teoretycznego podejścia do problemu zmniejszania ryzyka strat finansowych na podstawie transakcji terminowych, a zilustruję go danymi dotyczącymi transakcji terminowych na dolara amerykańskiego na WGT SA.

Zabezpieczanie i spekulacja

Zmniejszenia ryzyka strat finansowych można dokonywać na różne sposoby: można tworzyć portfele różnych opcji kupna i sprzedaży, można łączyć zakupy opcji wraz z wystawieniem opcji, a także zawarciem kontraktów terminowych (kontrakty commodity futures, dotyczące towarów, czy kontrakty financial futures, dotyczące instrumentów finansowych). Można też zestawiać

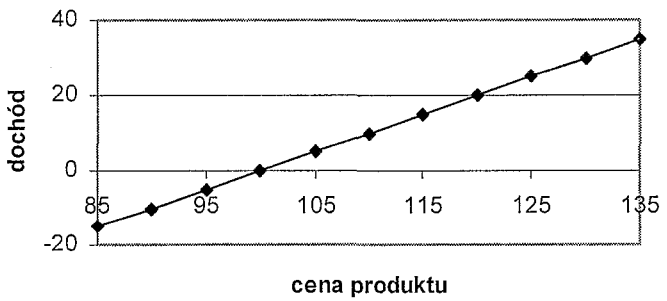
pochodne papiery wartościowe z podstawowymi czy, ogólniej, z produktem podstawowym. Takie strategie zabezpieczenia się przed niepożądanymi ruchami cen nazywane są hedgingiem albo asekuracją.

Hedging stosuje się przy planowaniu zakupu czy sprzedaży w przyszłości. Często podobne zachowania mają jednak charakter tylko spekulacyjny. Wielu autorów podkreśla, że spekulacja nie ma charakteru negatywnego, a często dzięki spekulantom właśnie osiąga się pożądaną płynność rynku.

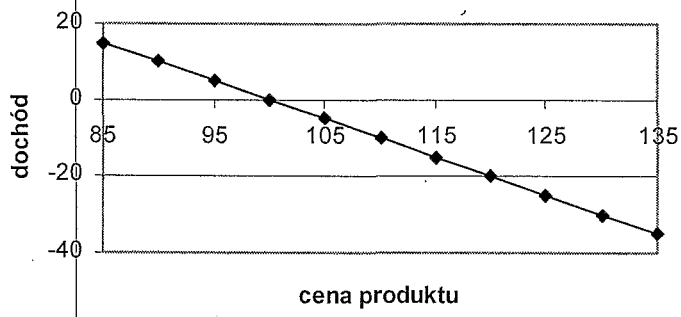
Zabezpieczanie na podstawie portfeli zawierających opcje zostało opisane w artykule [2]. Zastosowanie strategii z użyciem opcji zilustrowałem tam danymi z WGT SA z 1998 r. Mało rozwinięty rynek opcji na WGT SA jeszcze bardziej podupadł w 1999 roku na korzyść rynku kontraktów terminowych, które można zawierać od początku 1999 roku (pierwsze kontrakty zawarto 20 stycznia 1999 r.).

Kontrakty terminowe w zabezpieczaniu i spekulacji

Zacznijmy od dwóch bazowych strategii z użyciem transakcji terminowych. Nazywają się one „długi hedging" i „krótki hedging". Pierwsza to kupno kontraktu futures (czyli umowa dotycząca kupna produktu po określonej cenie), druga – jego sprzedaż (czyli zobowiązanie dotyczące sprzedaży produktu po umówionej cenie). Pierwsza zabezpiecza inwestora przed ewentualnym wzrostem ceny produktu, który zamierza on kupić za jakiś czas (rys. 1), druga – przed spadkiem ceny produktu, który posiada inwestor (rys. 2). Na rysunkach przedstawiono wykresy dochodu inwestora w obu przypadkach, przy umówionej w kontrakcie cenie równej 100 jednostek pieniężnych.



Rysunek 1
Zakup kontraktu terminowego



Rysunek 2
Sprzedaż kontraktu terminowego

Zupełnie podobnie kształtuje się dochód inwestora: w pierwszym przypadku przy kupnie produktu (jeśli można go przechowywać), w drugim przypadku przy pożyczeniu produktu (inwestor liczy, że przy oddawaniu ceny będą niższe). Rozważmy teraz dwa rodzaje zachowań na rynku: zabezpieczanie i spekulację w zastosowaniu do kontraktów terminowych.

Jak można przeczytać w [4], „zabezpieczanie polega na tym, że przyjmuje się pewną strategię, dzięki której pozycja zajmowana przez inwestora na rynku, będąca pozycją narażoną na ryzyko, zostaje przekształcona w pozycję częściowo lub całkowicie wolną od ryzyka". Weźmy przykład eksportera półtusz wieprzowych, który za swój towar ma otrzymać za trzy miesiące 10 000 dolarów USA. Narażony on jest na ryzyko spadku kursu dolara względem złotówki. Może on teraz zabezpieczyć się, zawierając transakcję terminową na sprzedaż 10 000 USD za trzy miesiące, powiedzmy w cenie 4,2 zł za dolara. Gdy cena dolara spadnie poniżej 4,2 zł, cena otrzymanych dolarów na rynku gotówkowym spadnie, ale będzie zrekompensowana dochodem ze sprzedanego kontraktu futures, a mówiąc prościej – eksporter zapewnił sobie sprzedaż dolarów po satysfakcjonującej go cenie. W przypadku wzrostu ceny dolara eksporter musząc wywiązać się z transakcji terminowej pozbawia się dodatkowych zysków, ale jego głównym celem było uwolnienie się od ryzyka, a nie spekulacja walutowa.

Podobny przykład można podać dla sytuacji, w której importer soi musząc zapłacić za np. cztery miesiące sumę 20 000 USD zabezpiecza się przed wzrostem kursu dolara w stosunku do złotówki, kupując kontrakt terminowy na 20 000 USD po np. 4,1 zł za dolara. Można powiedzieć, że złożenie „pozycji" inwestora na rynku gotówkowym z „pozycją" na rynku futures pozwala uniknąć strat, jak również i dodatkowych zysków.

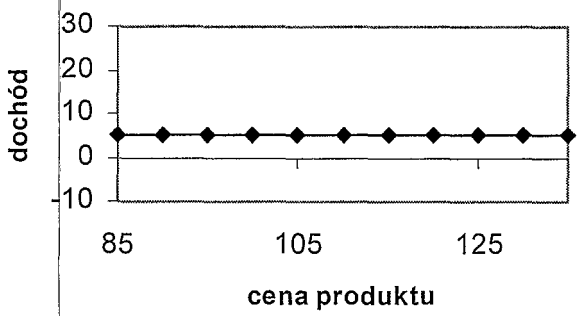
Co do spekulacji, to w [4] można przeczytać, że „spekulacja jest to działanie inwestora wynikające z akceptowanych przez niego prognoz, zwłaszcza prognoz cen instrumentów finansowych. Przy spekulacji liczy się na uzyskanie ponadprzeciętnego dochodu, akceptując jednocześnie często ponadprzeciętne ryzyko".

Najprostszym przykładem spekulacji byłoby np. zawarcie transakcji futures na kupno 10 000 USD od eksportera półtusz wieprzowych za trzy miesiące po 4,2 zł za dolara. Oczywiście spekulant liczy na wzrost kursu dolara w stosunku do złotówki. W takiej sytuacji będzie miał zyski. Wystawia się jednak na duże ryzyko spadku ceny dolara poniżej 4,2 zł. We wcześniejszym przykładzie importera soi mógł on np. zawrzeć swój kontrakt terminowy właśnie ze spekulantem, który liczy na spadek kursu dolara. Spekulant mógłby np. zestawić dwa kontrakty terminowe ze sobą: kupić kontrakt terminowy po niższej cenie dostawy, a sprzedać kontrakt na ten sam produkt po wyższej cenie dostawy (zawarcie takich transakcji w tym samym czasie jest, jak łatwo sobie wyobrazić, raczej rzadkim zdarzeniem).

Można powiedzieć, że zyski spekulantów czy zabezpieczających się są jednocześnie stratami innych uczestników rynku i odwrotnie: ich straty – zyskami innych uczestników rynku. Jedni szukają ryzyka, inni chcą się go ustrzec.

Portfele zawierające opcje

Wśród strategii służących czy to do zabezpieczania się przed niepożądanymi ruchami cen, czy to do spekulacji istnieje też wiele, w których zestawia się kontrakty terminowe z opcjami (obok portfeli złożonych z samych opcji, por. [2]). Z powodu mało rozwiniętego rynku opcji na WGT SA ograniczę się jedynie do zilustrowania takich zestawień jednym przykładem. Strategia „conversion" polega na zakupie kontraktu terminowego z niższą ceną rozliczenia w zestawieniu z kupnem opcji sprzedaży i wystawieniem opcji kupna z wyższymi cenami wykonania. Termin realizacji transakcji terminowej i terminy wygaśnięcia opcji są takie same i oczywiście wszystkie umowy dotyczą identycznego, w identycznej ilości produktu. Załóżmy, że cena w kontrakcie terminowym opiewa na 100 jednostek pieniężnych, opcja sprzedaży na cenę 110 jednostek kosztuje 10 jednostek, a opcję kupna wystawiono na 110 jednostek z premią opcyjną 5 jednostek. Można sprawdzić, że wykres wynikowy dochodu inwestora jest wykresem funkcji stałej (rys. 3).



Rysunek 3
Dochód w strategii „conversion"

W tabeli 1 zamieszczono składniki obliczeń.

Tabela 1

Cena produktu	Dochód z kontraktu terminowego	Dochód z opcji sprzedaży	Dochód z opcji kupna	Dochód sumaryczny
85	–15	15	5	5
90	–10	10	5	5
95	–5	5	5	5
100	0	0	5	5
105	5	–5	5	5
110	10	–10	5	5
115	15	–10	0	5
120	20	–10	–5	5
125	25	–10	–10	5
130	30	–10	–15	5
135	35	–10	–20	5

Oczywiście, nie zawsze otrzymamy wykres funkcji stałej, możliwych bowiem kombinacji, w których występują kontrakty terminowe i opcje, jest bardzo wiele.

WGT SA zamierza wprowadzić do obrotu opcje na kontrakty terminowe, jej plany są optymistyczne.

Ceny futures

Istotną możliwością stwarzaną przez giełdy terminowe jest możliwość wycofania się z kontraktu przed ustalonym w nim terminem dostawy. Na kolejnych sesjach giełdowych można „zamknąć swoją pozycję na rynku futures" zawierając kontrakt przeciwny. W przypadku kupionego kontraktu należy identyczny kontrakt sprzedać, w przypadku kontraktu sprzedanego – należy analogiczny kontrakt kupić. Oczywiście, zależy to jeszcze od możliwości znalezienia strony przeciwnej takiej transakcji. Stąd tak ważna jest kwestia płynności rynku. Jednak cena na produkt, którego kontrakt dotyczy, zmienia się. Pociąga to za sobą zmianę ceny ustalanej w kontraktach. Na zakończenie sesji giełda ustala „cenę zamknięcia", która jest pewną wypadkową cen w zawartych podczas sesji transakcjach na dany produkt o ustalonym terminie realizacji. Te wypadkowe zmieniających się cen na kontrakty futures, ustalane przez giełdę jako ceny zamknięcia sesji, nazywają się cenami futures albo cenami terminowymi. Okazuje się, że większość kontraktów terminowych zawieranych na giełdach towarowych jest zamykana przed terminem dostawy. W przypadku kontraktów na np. stopy procentowe fizyczne dostarczenie instrumentu bazowego w ogóle nie wchodzi w grę. W przypadku np. kontraktów na kurs określonej waluty nie chodzi o dostarczenie tej waluty, ale o rozliczenie zysku/straty obliczonych na podstawie różnicy ceny zapisanej w kontrakcie i aktualnej ceny terminowej w momencie zamykania kontraktu. Należy się spodziewać, że ta cena terminowa pokryje się z ceną gotówkową w terminie dostawy kontraktu (określonej w umowie). Pomijam tu kwestie związane z depozytami początkowymi, stanowiącymi gwarancję wywiązania się z umów uczestników rynku, i technicznymi sposobami pozwalającymi uprościć wiele działań uczestników rynku terminowego, w szczególności sprawę rozliczania na bieżąco.

Delta hedging. Współczynnik zabezpieczenia

Ciekawym typem strategii zabezpieczania się jest strategia nazywana delta hedgingiem. Tutaj postępowanie polega na sprzedaży lub zakupie produktu i utworzeniu pewnego pakietu opcji czy kontraktów terminowych, związanych z tym produktem bazowym. W przypadku giełdy towarowej strategia delta hedging mogłaby być zastosowana w odniesieniu do towaru nie psującego się,

który można spokojnie przechowywać, lub waluty. Strategia wymagałaby zaobserwowania wartości współczynnika delta, który w tym przypadku oznaczałby stosunek zmiany ceny futures do zmiany ceny odpowiadającego jej produktu. Należy się spodziewać, że współczynnik delta będzie w przybliżeniu stały, przy odległym terminie realizacji kontraktu w pewnym, niezbyt długim odcinku czasu. Strategia delta hedging ma na celu otrzymanie zestawu instrumentów niewrażliwego na zmiany cen instrumentu podstawowego. Mogłaby być np. zastosowana następująco:

- sprzedaż kontraktu terminowego i kupno towaru w proporcji 1 : delta,
- sprzedaż towaru (planowana) i sprzedaż kontraktu terminowego w proporcji delta : 1,
- kupno towaru (planowane) i zakup kontraktu terminowego w proporcji delta : 1.

Można powiedzieć, że w pierwszym przypadku zabezpieczyliśmy się przed wzrostem cen (straty w przypadku rozliczenia kontraktu niwelują się dzięki wzrostowi cen kupionego towaru), w drugim – przed spadkiem cen (straty na wartości sprzedawanego towaru niwelują się dzięki wzrostowi zysku na kontrakcie terminowym), w trzecim – przed wzrostem cen (straty z powodu wyższej ceny, jaką trzeba będzie zapłacić za kupowany towar, niwelują się dzięki wzrostowi zysku na kontrakcie terminowym). Strategia delta hedging pozwala nam zapewnić ustalony sumaryczny dochód, który również pozostanie w przybliżeniu stały przy ruchu cen w odwrotną stronę.

Zilustrujmy pierwsze zestawienie (sprzedaż kontraktu i kupno towaru) przykładem liczbowym. Załóżmy, że kontrakt zawarty w styczniu opiewa na sprzedaż 100 t pszenicy po 500 zł za tonę we wrześniu, a w lutym i marcu zaobserwowano, że wraz ze wzrostem ceny pszenicy na rynku rosła i cena futures w proporcji ok. 1 : 0,33. Współczynnik delta wyniósł więc ok. 0,33. Inwestor kupił pszenicę w ilości 33 t. Oznaczmy przez X cenę rynkową pszenicy, a przez Y cenę futures 1 kwietnia. Gdyby współczynnik delta się nie zmieniał, to przy nowej cenie rynkowej 15 kwietnia $X' = X + 1$ nowa cena futures wyniosłaby: $Y' = Y + 0{,}33$. Sumaryczny dochód inwestora wyniósł przed zwyżką cen: $Z = 100 \times (500 - Y) + 33 \times X$. Dochód po zwyżce cen wynosi: $Z' = 100 \times (500 - Y') + 33 \times X'$. Łatwo zauważyć, że $Z = Z'$, a więc dochód sumaryczny się nie zmienił.

Współczynnik delta jest stosowany w analogiczny sposób w przypadku opcji. Dla kontraktów terminowych definiuje się jeszcze wielkość zwaną współczynnikiem zabezpieczenia. Jego rola jest taka, jak współczynnika delta, niemniej sposób szacowania jest dokładniej opisywany. Współczynnik zabez-

pieczenia stanowi iloraz wielkości pozycji zajętej w kontraktach futures do wielkości pozycji zabezpieczanej.

Nie zawsze wartość jeden jest optymalna. Dowodzi się, że optymalna wielkość współczynnika zabezpieczenia (H*) wynosi:

$$H^* = r \times S_S/S_F$$

We wzorze występuje oszacowanie współczynnika korelacji (r) między delta F i delta S, ocena odchylenia standardowego delta S (S_S) oraz odchylenia standardowego delta F (S_F);
delta S – są to zmiany ceny rynkowej instrumentu bazowego, delta F – zmiany ceny terminowej. Te wartości dotyczą okresu stosowania transakcji zabezpieczającej.

Wartość H* minimalizuje wariancję całkowitego dochodu inwestora. Całkowity dochód jest rozumiany jako suma dochodów pochodzących od aktywu bazowego i zysków/strat pochodzących od kontraktów futures.

Oszacowania wartości występujących we wzorze na H* dokonuje się na podstawie delta S i delta F obliczonych dla rozłącznych przedziałów czasowych wypełniających okres trwania transakcji. Warto zauważyć, że obliczone H* jest to współczynnik regresji liniowej delta S względem delta F.

Postępowanie na WGT i podsumowanie

Na zakończenie zobaczmy, jak mógłby postąpić eksporter półtusz wieprzowych, który np. 8 marca 2000 r. rozważał możliwość zawarcia kontraktu terminowego na sprzedaż dolarów amerykańskich. Otrzyma on 10 000 USD za swój towar 7 czerwca 2000 r. i natychmiast będzie chciał wymienić dolary na złotówki. Jeśli obawia się spadku ceny dolara, która 8 marca wynosiła 4,1312 zł, to zapewne cena futures na kontrakt terminowy na dolara na WGT SA z tego samego dnia, wynosząca 4,295 zł/USD, wyda mu się atrakcyjna. Na WGT SA jednostką transakcyjną jest właśnie 10 000 USD, więc eksporter mógłby sprzedać jeden kontrakt terminowy na czerwiec (jeśli znajdzie chętnego kupującego), czyli zastosować krótki hedging. Chcąc zrealizować swój kontrakt 8 czerwca, przekona się, że stracił na kontrakcie, gdyż cena dolara osiągnęła wtedy wartość 4,396 zł.

Gdyby eksporter dopuszczał możliwość późniejszej sprzedaży dolarów, to mógłby zawrzeć kontrakt nie na czerwiec, lecz na późniejszy miesiąc, a zawsze mógłby próbować „zamknąć swoją pozycję" na rynku futures już wcześniej,

np. 8 czerwca, gdy uzna, że nie warto czekać. Jest to związane z możliwością zawarcia kontraktu przeciwnego w czasie trwania kontraktu i rozliczenia się po aktualnych cenach futures. Hull [1] pisze, że „Z reguły (...) wybierany jest kontrakt o późniejszym terminie dostawy. Jest to związane z faktem, że w niektórych wypadkach ceny terminowe w okresie dostawy mogą być bardzo niestabilne".

Weźmy teraz przypadek importera soi, który w połowie listopada 1999 r. dowiaduje się, że za import soi będzie musiał w połowie marca 2000 r. zapłacić 20 000 USD. Aby zabezpieczyć się przed wzrostem ceny dolarów, które zamierza kupić w ostatniej chwili, może zawrzeć dwa kontrakty na kupno dolarów na WGT SA (jeśli znajdzie chętnego sprzedającego), czyli zastosować długi hedging. 17 listopada 1999 cena zamknięcia na kontrakt marcowy wynosiła wprawdzie 4,334 zł/USD, co było większą sumą niż na rynku (cena rynkowa była równa 4,242 zł/USD), ale importer liczy na większą zwyżkę do marca. Ponadto, nie płacąc w tej chwili obniża faktyczne koszty zakupu dolarów (np. możliwość oprocentowania złotówek na rachunku bankowym). W tym przypadku zabezpieczenie okazuje się nieopłacalne, gdyż cena rynkowa 15 marca 2000 r. wyniosła 4,082 zł/USD. Importer może jednak zawrzeć kontrakty przeciwne dla zamknięcia swojej pozycji na rynku futures w okresie listopad – marzec, gdy uzna, że obserwowany malejący trend kursu dolara nie zmieni się do połowy marca. Wtedy jego straty będą niższe. Można było zaobserwować, że ceny futures na kurs dolara amerykańskiego na marzec 2000 r. po 17 listopada 1999 r. nieco spadły, by wzrosnąć w początku grudnia (np. 2 grudnia wyniosły 4,353 zł/USD). Potem ceny znowu raczej spadały i np. 4 stycznia 2000 r. wyniosły 4,18 zł/USD. Można sobie wyobrazić, że np. wtedy inwestor mógł zawrzeć kontrakty przeciwne, czyli na sprzedaż dolarów (dwa kontrakty po 10 000 USD) zamykając w ten sposób swoje pozycje na rynku futures i mógł postanowić, że kupi jeszcze raz takie kontrakty, gdy tylko cena futures osiągnie co najmniej 4,3 zł/USD. W takiej sytuacji nie zrealizowałby tego kupna, gdyż ceny wahając się spadały i do 15 marca 2000 r. osiągały wartości niższe (najwyższa wartość wyniosła 2 lutego 2000 r. 4,279 zł/USD). Straty importera wyniosłyby w tym przykładzie $20\,000 \times (4{,}334 - 4{,}18)$, co jest sumą mniejszą niż na początku, gdy ta wielkość byłaby równa $20\,000 \times (4{,}334 - 4{,}082)$.

Gdyby importer chciał zdecydować się na kupno kontraktu o późniejszym terminie dostawy (o korzystności takiego rozwiązania wspomniano wcześniej), mógłby jeszcze czekać licząc na zyski przy zwyżce ceny dolara. Np. kurs dolara USA wynosił 5 maja 2000 r. 4,6342 zł, cena terminowa dla kontraktu na czerwiec 2000 r. wynosiła 4,718 zł. W listopadzie 1999 r. realne było kupienie kontraktu na dolara USA na czerwiec 2000 r. Niestety, mimo formalnego roz-

poczęcia obrotu kontraktami na czerwiec 2000 r. już w lipcu 1999 r., pierwsze kontrakty zawarto dopiero w styczniu 2000 r. Nie ma zatem danych o cenie takiego kontraktu w listopadzie 1999 r. Należy jednak przypuszczać, że ta cena, gdyby importer złożył ofertę kupna, ustaliłaby się na poziomie wprawdzie wyższym od ceny zamknięcia na kontrakt marcowy, która, jak podano wcześniej, w dniu 17 listopada wynosiła 4,334 zł/USD, ale zdecydowanie niższym niż 4,718 zł/USD. Importer mógłby więc zyskać na kontrakcie.

Jak pokazano, oprócz podjęcia dobrej decyzji na początku należy jeszcze i później bacznie obserwować zachowania na giełdzie.

Co do strategii delta hedging czy też obliczania współczynnika zabezpieczenia, to zobaczmy, jaka powinna być wielkość kontraktu futures na sprzedaż dolara amerykańskiego w czerwcu 2000 r. w przypadku eksportera półtusz wieprzowych.

Oszacowano współczynnik zabezpieczenia dla kontraktu na dolara na czerwiec 2000 r. na podstawie 10-dniowych przedziałów czasowych jako bazy dla znalezienia delta S i delta F. Wartość współczynnika wynosi 0,959. Ciekawe, że dla okresów jednodniowych wynik wyszedł podobny i równy 0,949. Obydwie wartości trzeba jednak zaokrąglić do liczby całkowitej. Zastosowany współczynnik równy jeden był więc dobrym rozwiązaniem. Współczynnik delta daje zdecydowanie różne wartości w zależności od długości okresów, dla których oblicza się przyrosty ceny instrumentu bazowego i kontraktu terminowego. Dla okresów jednodniowych jest on bardzo zróżnicowany, a jego średnia wartość daleka od jedynki. Dla przedziałów czasowych dłuższych średnia wartość zbliża się do jedynki i np. dla okresów 10-dniowych wynosi 1,048 (co w przybliżeniu daje odwrotność współczynnika zabezpieczenia dla takich samych okresów).

Można chyba przypuszczać, że dla większości kontraktów na dolara amerykańskiego proporcje między wielkością pozycji zajętej w kontraktach terminowych a wielkością pozycji zabezpieczanej wynosić powinny 1 : 1.

Co do spekulacji, to można sobie wyobrazić sytuację, w której to właśnie spekulant sprzedał importerowi soi dwa kontrakty na dolara USA na marzec 2000 r. 17 listopada 1999 r. Potem też on mógł kupić te dwa kontrakty 4 stycznia 2000 r. Zatem kupił taniej (po 4,18 zł), sprzedał drożej (po 4,334 zł), tyle że nie w tym samym czasie. Cała strata importera (pomijając prowizje maklerskie) przeszła w jego ręce.

Literatura

1. HULL J., 1998: *Kontrakty terminowe i opcje*. Wprowadzenie, WIG-Press, Warszawa.
2. JABŁONOWSKI S., 2000: Strategie zmniejszania ryzyka strat finansowych na Warszawskiej Giełdzie Towarowej, *Zagadnienia Ekonomiki Rolnej*, 1 (276).
3. JAJUGA K., JAJUGA T., 1994: *Jak inwestować w papiery wartościowe*, Wydawnictwo Naukowe PWN, Warszawa.
4. JAJUGA K., KUZIAK K., MARKOWSKI P., 1997: *Inwestycje finansowe*, Wydawnictwo AE im. O. Langego, Wrocław.
5. JANUSZKIEWICZ W. (red.), 1991: *Giełdy w gospodarce światowej*, PWE, Warszawa.
6. WGT SA, *Archiwalne dane dotyczące statystyki handlu...*, WWW.WGT.COM.PL
7. *Średnie kursy NBP*, http://yogi.pl/gielda/waluta.usd.html.

Futures contracts in strategies of hedging against unprofitable prices' changes on Warsaw Board of Trade

Abstract

I the paper, chosen investing strategies contracted on boards of trade are described. The aim of these strategies is to hedge an investor against prices' changes undesirable to him. The strategies are illustrated by data from Warsaw Board of Trade Co from 1999 and 2000. Among numerous possible hedgings against undesirable prices' changes of commodities, currencies' courses and rates, investings with futures contracts are marked out. Futures contracts have been developped on Warsaw Board of Trade since the beginning of 1999. The options' market was weak and declined since the end of 1998. Probably, it will not develop more. Futures transactions on currencies' courses and rates are of the biggest significance because of commodity futures have been contracted very rarely.

The described chosen strategies take into cosideration purchase and sale various futures contracts and combinations of purchase and sale contracts with basic instrument's purchase or sale. The strategies have been analysed on data from Warsaw Board of Trade.

Arkadiusz Orłowski
Katedra Ekonometrii i Informatyki SGGW
Mariusz Bęben
Szkoła Nauk Ścisłych w Warszawie

Metody chaosu deterministycznego w badaniach ekonomicznych

Wstęp

Chaos deterministyczny jest teorią usiłującą wyjaśnić nieregularne, podobne do stochastycznych, własności i zachowanie się części nieliniowych układów deterministycznych. Teoria ta przyciąga uwagę coraz większej liczby naukowców z różnych dziedzin wiedzy, w tym ekonomistów i finansistów. O układzie powiemy, że jest czysto deterministyczny, jeśli jego opis nie wymaga wprowadzenia zmiennych losowych. W przypadku, gdy obserwowany system umiemy opisać jedynie przez podanie prawdopodobieństw przejść pomiędzy możliwymi stanami układu, mówimy, że system jest stochastyczny.

Ważną cechą zachowania układu chaotycznego jest ograniczona możliwość tworzenia prognoz długoterminowych, co wiąże się z dużą wrażliwością układu na warunki początkowe. Dowolne trajektorie startujące z podobnych warunków początkowych będą się szybko oddalać od siebie: odległość między nimi będzie rosła wykładniczo z czasem.

Zanim przejdziemy do omówienia szczegółów natury technicznej, warto jeszcze wspomnieć o uzasadnieniu zastosowania tych technik do danych pochodzących z obserwacji systemu, o którym nie wiadomo dokładnie, czy jest całkowicie deterministyczny. Odpowiedź na tak postawione pytanie brzmi: zastosowanie teorii chaosu w takim przypadku może być kontrowersyjne. Teoria chaosu deterministycznego tłumaczy zachowania systemów o znanej deterministycznej dynamice nieliniowej, opisanej przez zależności funkcyjne typu: $x_{n+1} = f(x_n)$, gdzie x_n jest wektorem opisującym stan układu w skończenie wymiarowej przestrzeni fazowej w chwili n, a f jest funkcją opisującą ewolucję czasową układu. W przypadku danych eksperymentalnych mamy zazwyczaj do dyspozycji jedynie skalarny szereg czasowy obserwacji będących pewną nieznaną funkcją wektorów x: $s_n = s(x_n)$. Chcemy więc wierzyć, że te obserwa-

cje pozwolą nam zrekonstruować przestrzeń fazową oraz że własności pewnych wielkości w przestrzeni zrekonstruowanej będą takie same jak w pierwotnej, do czego daje podstawy omówiona później technika zanurzeń.

W tej sytuacji możliwych jest kilka podejść. Pierwsze z nich, rygorystyczne, nakazywałoby najpierw zweryfikować tezę o determinizmie i nieliniowości systemu. Jednak trudno jest o prosty dowód niskowymiarowej chaotyczności danego systemu. Jego skonstruowanie wymagałoby zaobserwowania samopodobieństwa w danych pomiarowych – czyli, inaczej mówiąc, ich skalowania się (taka próba została przez nas podjęta dla rozkładów dziennych stóp zwrotu). Poza tym, należałoby pokazać wrażliwość naszego układu na warunki początkowe, co również usiłujemy przeprowadzić, obliczając maksymalny wykładnik Lyapunova. W przypadku idealnym powinniśmy umieć wyekstrahować z danych postać funkcji f opisującej dynamikę układu, za jej pomocą obliczyć wartości przyszłe obserwacji i sprawdzić, czy mają one własności statystyczne analogiczne z pierwotnymi danymi. Takie podejście jest, z wielu powodów, bardzo trudne do zrealizowania i jego rygorystyczne przestrzeganie z pewnością doprowadziłoby do zarzucenia stosowania metod nieliniowych w dziedzinach, w których okazały się one przydatne. Wielu autorów przyjmuje więc bardziej pragmatyczny paradygmat i nie zakładając nic na temat dynamiki opisującej układ, koncentruje się na pytaniu nie o obecność w nim chaosu deterministycznego, ale o to, czy metody wypracowane przy badaniu nieliniowych modeli teoretycznych o znanej dynamice są przydatnym językiem do opisu posiadanych danych. Czysty determinizm w systemach rzeczywistych jest bardzo trudny do zaobserwowania, ponieważ wszystkie te systemy oddziałują z otoczeniem. Tak więc obraz deterministyczny jest raczej przypadkiem granicznym. Na domiar złego, nawet systemy, o których wiemy z pewnością, że są nieliniowe mogą wykazywać zachowania stochastyczne wskutek uśrednienia do rozkładu gaussowskiego wielu słabo ze sobą oddziałujących stopni swobody.

Nieliniowa analiza szeregów czasowych ma swój początek w latach 80. dwudziestego wieku i opiera się na twierdzeniu o zanurzeniu (twierdzenie Takensa), które w dalszej części pracy przytaczamy w całości. Jej rozwój wywodzi się z odkrycia osobliwych chaotycznych atraktorów w nieliniowych systemach dysypatywnych. Systemy dysypatywne odznaczają się obecnością sił tarcia w przeciwieństwie do systemów zachowawczych, czy też inaczej hamiltonowskich, które najczęściej są przedmiotem analizy w fizyce teoretycznej. W systemach dysypatywnych nie funkcjonują proste zasady zachowania energii, pędu, momentu pędu. Poglądowo mówiąc, objętości zajmowane przez trajektorie w przestrzeni fazowej kurczą się, a same trajektorie zbiegają do pewnych atraktorów. W systemach nieliniowych mogą występować niestabilne

punkty stałe, czyli, inaczej mówiąc, kierunki lokalnej ekspansji, ale pomimo tego system jako całość pozostaje dysypatywny, czyli nie opuszcza ustalonego obszaru w przestrzeni fazowej. Powyższe cechy powodują występowanie tzw. dziwnych atraktorów, które często bywają fraktalami.

W dalszej części przedstawimy teoretyczne podstawy techniki zanurzenia, która bazuje na wspomnianym już wcześniej twierdzeniu Takensa, jak również zaprezentujemy metodę surogatów, za pomocą której można wygenerować szereg czasowy o takich samych własnościach charakteryzujących procesy liniowe, jak wyjściowe, aby następnie sprawdzić, czy tak wyprodukowany surogat różni się pod względem charakterystyk nieliniowości od oryginału, co by było sygnałem, że dane wyjściowe nie są realizacją liniowego procesu.

Technika zanurzeniowa

Jak wcześniej wspomnieliśmy, większość obserwacji dokonywanych na systemach jest skalarna, podczas gdy stan systemu opisywany jest wektorem w d-wymiarowej przestrzeni fazowej. Niech **x**(t) będzie d-wymiarowym wektorem w przestrzeni fazowej, a s(**x**(t)) skalarną obserwacją dokonaną na systemie w chwili t. Metoda zanurzeniowa opiera się na założeniu, że wartości przeszłe i przyszłe skalarnej obserwacji zawierają informację na temat bieżących wartości nie obserwowanych stopni swobody. Wektorem w przestrzeni zanurzeniowej (embadding space) o wymiarze E i przesunięciu czasowym (time delay) τ nazwiemy wektor o współrzędnych (s(t-(E-1)τ), s(t-(E-2)τ), ..., s(t)). Według przedstawionego poniżej twierdzenia Takensa, przestrzeń zanurzeniowa skonstruowana według pewnych zasad rekonstruuje pierwotną przestrzeń fazową.

Twierdzenie o zanurzeniu Takensa (1981). *Niech d-wymiarowy system dynamiczny* $\dot{\vec{x}} = F(\vec{x})$, *bądź alternatywnie* $\vec{x}_{n+1} = f(\vec{x}_n)$, *posiada atraktor o wymiarze fraktalnym D < d. Na atraktorze dokonuje się pomiarów skalarnych* $s(t) = s(\vec{x}(t))$, *a następnie konstruuje się przestrzeń zanurzeniową (s(t-(E-1)*τ*), s(t-(E-2)*τ*), ..., s(t)) o wymiarze E i przesunięciu czasowym* τ. *Wówczas topologia atraktora w przestrzeni zanurzeniowej o wymiarze* $E \geq 2D + 1$ *jest identyczna z topologią atraktora w przestrzeni fazowej* $\vec{x}(t)$ *teoretycznie dla wszystkich* τ *przy nieskończenie długim i dokładnym pomiarze s(t). W szczególności niezmienniki topologiczne takie jak wykładniki Lyapunova czy wymiary fraktalne mają te same wartości liczbowe w przestrzeni zanurzeniowej, co w oryginalnej przestrzeni fazowej.*

Ponieważ liczba stopni swobody w oryginalnym systemie nie jest znana, nie potrafimy podać wymiaru przestrzeni, którego powinniśmy użyć do zanurzenia. Oczywiście, zgodnie z twierdzeniem Takensa, jeśli weźmiemy dostatecznie wysoki wymiar zanurzenia, jesteśmy pewni, że uzyskamy rekonstrukcję atraktora o tych samych własnościach topologicznych, co oryginał. Przy obliczeniach zainteresowani jesteśmy jednak znalezieniem najmniejszego możliwego wymiaru zanurzenia E ze względu na skończoność naszego szeregu pomiarów *s(t)* oraz rosnącą wraz ze wzrostem wymiaru złożonością obliczeń. Jedną z metod oszacowania najmniejszego koniecznego wymiaru zanurzenia E_{min} jest metoda fałszywych najbliższych sąsiadów (false nearest neighbours; Kennel, Brown, Abandel 1992). Metoda ta oparta jest na własności istnienia dyfeomorfizmu pomiędzy atraktorem w przestrzeni fazowej i jego rekonstrukcją poprzez zanurzenie. Zapewnia ono odwzorowywanie punktów z otoczenia x w punkty z otoczenia u, jeśli u jest odwzorowaniem x. W przypadku jednak, kiedy użyjemy zbyt niskiego wymiaru zanurzenia, czyli $E < E_{min}$, w otoczeniu punktu u znajdą się punkty, które nie należą do otoczenia punktu x na oryginalnym atraktorze w przestrzeni fazowej. Są to fałszywi sąsiedzi. Ich liczba zmniejsza się wraz ze zwiększaniem wymiaru zanurzenia, aż przy pewnej wartości $E = E_{min}$ osiąga wartość zero. Wymiar przestrzeni zanurzeniowej, przy którym liczba fałszywych sąsiadów wynosi zero po raz pierwszy, przyjmujemy jako minimalny wymiar zanurzenia.

Następnym problemem, na który natrafiamy przy rekonstrukcji atraktora w przestrzeni zanurzeniowej, jest określenie przesunięcia czasowego τ (time delay). Do określenia właściwego przesunięcia czasowego posłużymy się metodą zaprezentowaną po raz pierwszy przez Frasera i Swinneya (Fraser, Swinney 1986). Metoda ta posługuje się pojęciem informacji wzajemnej (mutual information), która jest pewnym uogólnieniem funkcji autokorelacji szeregu czasowego. Nasze zrozumienie tej metody jest następujące: badamy uogólnioną autokorelację dla różnych przesunięć czasowych τ. W miarę jak τ rośnie autokorelacja ta powinna się zmniejszać, aż osiągnie pewne minimum. Osiągnięcie minimum przy danym τ oznacza, że wartość szeregu czasowego τ jednostek czasowych do tyłu w najmniejszym jak do tej pory stopniu wpływa na wartość obecną. Przyjmujemy więc, że w tym minimum wartość jest najsilniej zależna od pozostałych stopni swobody – czyli jest jakąś funkcją nieznanych stopni swobody, która mówi nam w sposób pośredni o ich zachowaniu. Zatem obieramy za wartość przesunięcia czasowego takie τ, dla którego wartość informacji wzajemnej osiąga pierwsze minimum.

Wykładniki Lyapunova i metoda surogatów

Cechą charakterystyczną systemów posiadających dynamikę nieliniową jest ich wrażliwość na warunki początkowe. Należy to rozumieć w ten sposób, że początkowo bliskie sobie trajektorie w przestrzeni fazowej oddalają się od siebie wraz z upływem czasu, a ich odległość rośnie wykładniczo z czasem. Miarą szybkości oddalania się tych trajektorii są właśnie wykładniki Lyapunova. W naszej pracy posługujemy się następującą definicją maksymalnego wykładnika Lyapunova (Hegger, Kantz 1999). Niech $\vec{s}_n$ oznacza punkt na zrekonstruowanej trajektorii układu w przestrzeni fazowej. Weźmy pod uwagę punkt z pewnego otoczenia $\vec{s}_n$ i oznaczmy go $\vec{s}_{n'}$. Dalej niech $\Delta_0 = s_n - s_{n'}$ będzie odległością pomiędzy rozpatrywanymi punktami. Po l iteracjach odległość ta wyniesie $\Delta_l = s_{n+l} - s_{n'+l}$. Jeśli pomiędzy nimi zachodzi relacja $\Delta_l \approx \Delta_0 e^{\lambda \cdot l}$, to wtedy λ jest maksymalnym wykładnikiem Lyapunova dla systemu dynamicznego. Obliczanie estymatora nieobciążonego największego wykładnika przebiega następująco. Najpierw obliczamy:

$$S(\varepsilon, m, t) = \left\langle \ln\left(\frac{1}{|U_n|} \sum_{s_{n'} \in U_n} |s_{n+t} - s_{n'+t}| \right) \right\rangle_n ,$$

gdzie, U_n oznacza ε-otoczenie punktu $\vec{s}_n$, a m jest wymiarem zanurzenia. Jeśli $S(\varepsilon, m, t)$ wykazuje liniowy wzrost z czasem t z tym samym nachyleniem dla m większych niż pewna krytyczna jego wartość i dla pewnego zakresu ε, to nachylenie to może być wzięte za estymator szukanej wielkości.

Do tej pory podaliśmy opis metody analizy szeregów czasowych przy założeniu, że pochodzą one z układów charakteryzujących się dynamiką nieliniową i wykazują zachowania chaotyczne. Zazwyczaj jednak mając szereg danych pomiarowych nie potrafimy powiedzieć, z jakiego rodzaju układu one pochodzą, jaka dynamika nimi rządzi: stochastyczna czy deterministyczna. W szczególności pytanie to dotyczy układów, którymi się zajmujemy – ekonomicznych. Potrzebne jest więc narzędzie statystyczne, które pozwoli porównać dane z danymi wygenerowanymi przez proces stochastyczny, ale nie dowolny, lecz taki, który ma takie same lub bardzo podobne charakterystyki (momenty,

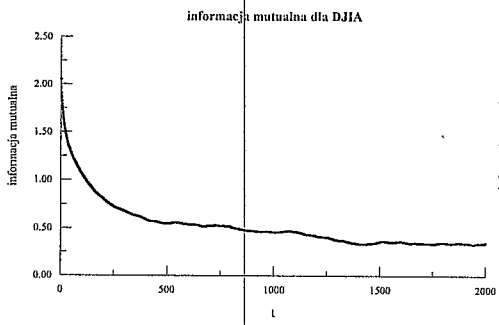
rozkład itd.) Do stworzenia takich szeregów czasowych przydatna jest metoda surogatów (Theiler, Eubank, Longtin, Galdrikian, Farmer 1992)

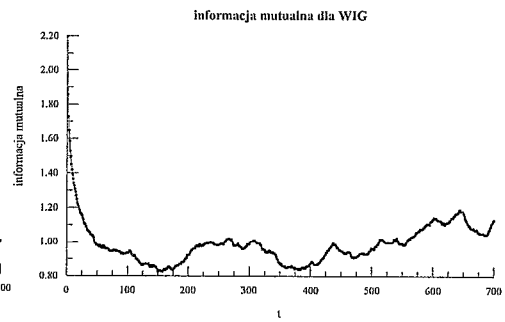
Surogat danego szeregu czasowego jest realizacją pewnego liniowego procesu stochastycznego, który ma niektóre cechy wspólne z pierwotnymi danymi. Cechami tymi są charakterystyki liniowe oraz rozkład prawdopodobieństwa wartości osiąganych przez szereg czasowy. Taki stan rzeczy jest osiągnięty poprzez zastosowanie dwóch twierdzeń. Pierwsze z nich mówi o relacji pomiędzy współczynnikami liniowego procesu stochastycznego a macierzą kowariancji (czy też funkcją autokorelacji), a mianowicie, że pierwsze jednoznacznie wyznaczają drugie i na odwrót. Relacja pomiędzy tymi dwoma charakterystykami jest znana pod nazwą równań Yule-Walkera. Krótko mówiąc, znając współczynniki procesu stochastycznego, potrafimy wyliczyć macierz kowariancji i na odwrót.

Drugie twierdzenie mówi o relacji pomiędzy widmem fourierowskim szeregu czasowego a funkcją autokorelacji tego szeregu (czy też macierzą kowariancji). Jest ono znane pod nazwą twierdzenia Wienera-Khinchina i mówi, że transformata Fouriera funkcji autokorelacji jest równa widmu fourierowskiemu szeregu czasowego. Pamiętamy przy tym, że obliczając widmo fourierowskie szeregu, tracimy informację o fazie transformaty Fouriera szeregu czasowego, której potrzebujemy, jeśli chcemy za pomocą odwrotnej transformaty odzyskać pierwotny szereg czasowy z widma. Fazę tę możemy więc dobrać dowolnie i nadal być pewnym, że odtworzony w ten sposób szereg czasowy będzie miał tę samą funkcję autokorelacji – gwarantuje to twierdzenie Wienera-Khinchina. Oczywiście, jeśli chcemy odtworzyć pierwotny szereg czasowy, musimy znać też pierwotne fazy – opisany wyżej sposób generowania szeregów czasowych (zwany randomizacją faz) pozwala uzyskać inny szereg czasowy o tej samej funkcji autokorelacji, a więc i o tych samych współczynnikach liniowego procesu stochastycznego, jeśli pierwotny szereg czasowy pochodzi z takiego procesu.

Badanie rzeczywistych szeregów czasowych

W pracy posługujemy się danymi pochodzącymi przede wszystkim z dwóch giełd papierów wartościowych: New York Stock Exchange (NYSE) i Warszawskiej Giełdy Papierów Wartościowych (WGPW). W przypadku NYSE bierzemy pod uwagę wskaźnik Dow Jones Industrial Avarage (DJIA) z okresu od 1930.03.04. do 2000.01.07 a dla giełdy warszawskiej posługujemy się Warszawskim Indeksem Giełdowym z okresu od 1991.04.16. do 2000.01.07.



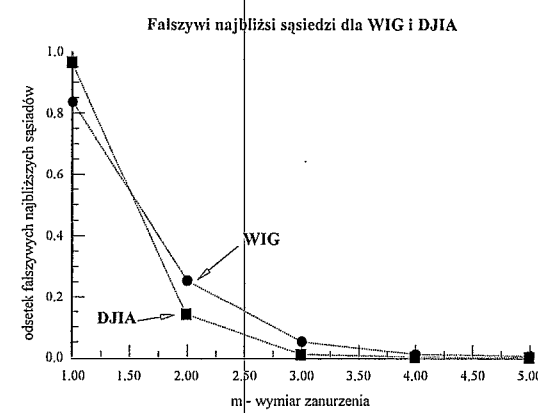


Aby zrekonstruować przestrzeń fazową, należy najpierw oszacować dwa parametry niezbędne do rekonstrukcji. Są nimi przesunięcie czasowe (time delay) τ oraz wymiar zanurzenia m. Pierwszą z tych wielkości określamy za pomocą informacji wzajemnej.

Na wykresach widać, że dla indeksu DJIA pierwsze ewentualne minimum pojawia się około $\tau = 1500$. Tę wartość przyjmujemy więc jako pierwszy parametr do rekonstrukcji przestrzeni fazowej DJIA. Jako przesunięcie czasowe dla WIG przyjmujemy wartość $\tau = 150$, ponieważ funkcja informacji wzajemnej dla tej wartości ma pierwsze wyraźne minimum.

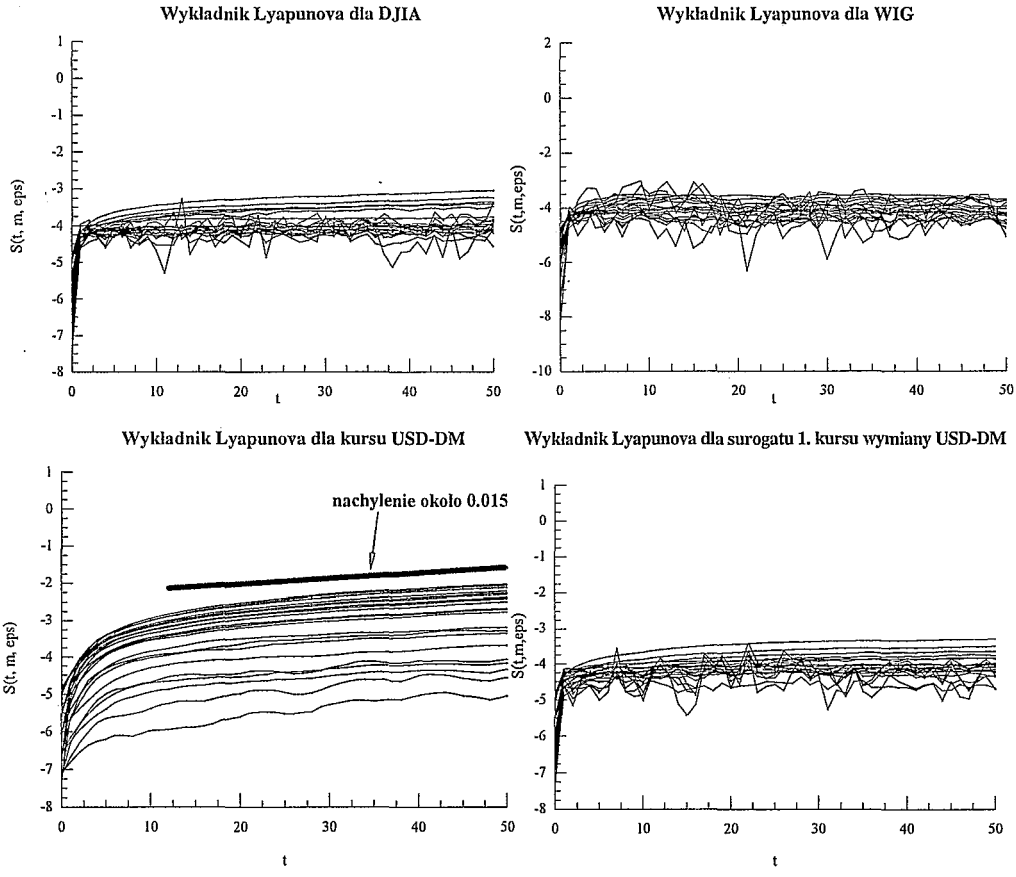
Do ustalenia minimalnego wymiaru zanurzenia posłużymy się wspomnianą już metodą fałszywych najbliższych sąsiadów. Wykres obok pokazuje odsetek fałszywych sąsiadów dla obu badanych indeksów jako funkcję wymiaru zanurzenia. Widzimy, że odsetek fałszywych najbliższych sąsiadów osiąga wartość bliską zera już dla m = 3. Tego wymiaru zanurzenia używamy dalej do naszych celów.

Odrębną klasą stosowanych przez nas danych są kursy wymiany walut. W naszej pracy posługujemy się kursem dolara amerykańskiego USD do marki niemieckiej DEM z lat 1971 do 2000 jako reprezentantem tej klasy danych. Przyjęte na podstawie obliczeń podobnych do tych przeprowadzonych dla indeksów giełdowych parametry rekonstrukcji przestrzeni fazowej to wymiar zanurzenia m = 3 oraz opóźnienie czasowe (time delay) $\tau = 500$. Z notowań USD-DEM również usunęliśmy trend wykładniczy.



Maksymalny wykładnik Lyapunova

Rysunki zamieszczone w tym rozdziale pokazują zależność S od czasu. Przedstawiliśmy powyższą zależność dla czterech przypadków: indeksów DJIA, WIG, kursu wymiany USD-DEM oraz dla surogatu kursu wymiany USD-DEM. Parametry użyte przez nas w poszczególnych przypadkach to: wymiar zanurzenia m dla wszystkich obrazów od m = 2 do m = 8. Podobnie liczba kroków iteracji dla wszystkich przypadków jest wspólna i wynosi 50. Rozmiar sąsiedztwa ε branego pod uwagę jest z przedziału od wartości szerokości przedziału danych podzielonych przez 1000 do tejże wartości podzielonej przez 100 i w poszczególnych przypadkach wynosi: dla USD-DEM od 8,34E-4 do 8,35E-3, dla DJIA od 2,17E-3 do 2,17E-2 i dla WIG od 3,19E-3 do 3,19E-2, gdzie E oznacza podnoszenie 10 do potęgi x.



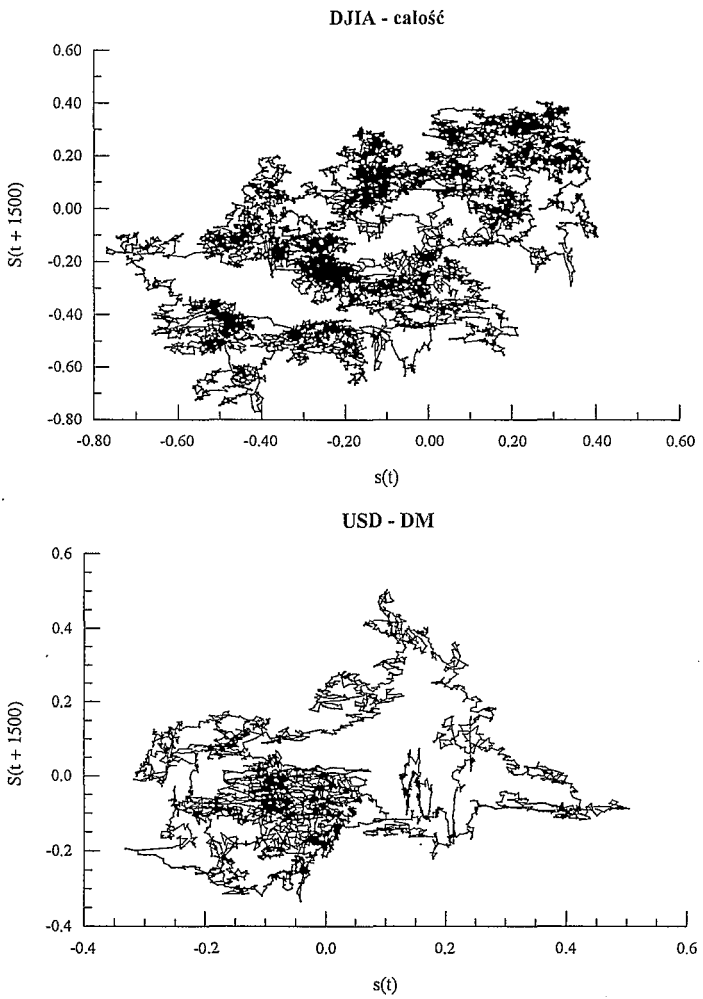
Dla obu indeksów giełdowych WIG oraz DJIA nie można dostrzec na rysunkach obszarów liniowości dla wystarczającej ilości wymiarów zanurzenia i rozmiarów sąsiedztwa (każda linia na rysunku oznacza jedną wartość m i ε). Oznacza to, że nie potrafimy oszacować wykładnika Lyapunova dla tych szeregów czasowych, bądź ze względu na duże zaszumienie danych, bądź wykładnik Lyapunova dla tych danych wynosi zero.

Inaczej rzeczy się mają w przypadku kursu wzajemnego walut USD-DEM. Na odpowiadającym mu wykresie można zaobserwować liniowy region wzrostu S z t. Prosta umieszczona nad krzywymi będącymi wykresami S została dopasowana do liniowego obszaru wzrostu, a jej współczynnik nachylenia wynosi około 0,015, co przyjmujemy również za szacowaną wartość wykładnika Lyapunova dla USD-DEM. Dla weryfikacji tezy o dodatniości wykładnika Lyapunova dla USD-DEM obliczyliśmy tę wielkość również dla surogatu kursu wymiany omawianych walut. Wynik obliczeń dla jednego z surogatów przedstawiony jest na rysunku po lewej stronie rysunku dla USD-DEM. Świadczy on dobitnie o tym, że kurs wymiany USD-DEM nie może być realizacją liniowego procesu stochastycznego, gdyż gdyby tak było, jego charakterystyka nieliniowa, którą w tym przypadku jest wykładnik Lyapunova, byłaby taka sama, jak dla jego surogatu. Identyczne wyniki daje porównanie omawianego kursu z kilkoma dalszymi swoimi surogatami. Znaleźliśmy zatem bardzo ważki dowód na to, że za notowaniami kursu walut USD-DEM kryje się nieliniowa dynamika deterministyczna, gdyż dodatni wykładnik Lyapunova jest jedną z zasadniczych cech chaosu deterministycznego.

Podsumowując: udało nam się dowieść istnienia dodatniego wykładnika Lyapunova dla kursu USD-DEM. Wynik ten skonfrontowaliśmy z obliczeniami przeprowadzonymi dla surogatów. Nie wykryliśmy natomiast dodatnich wykładników dla obu badanych przez nas indeksów giełdowych. O ile w pierwszym przypadku znaleźliśmy silny argument za obecnością chaosu deterministycznego w badanych danych, o tyle pozostałe przypadki zdają się świadczyć o czymś przeciwnym.

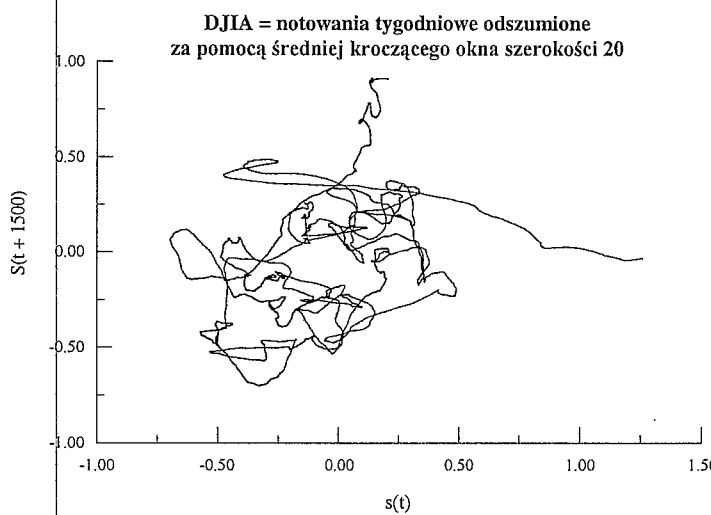
Rekonstrukcja przestrzeni fazowej

Rekonstrukcja przestrzeni fazowej, dla której podstawą jest cytowane przez nas we wstępie twierdzenie Takensa o zanurzeniu, umożliwia wizualizację dynamiki, której podlega badany system. W tym rozdziale przedstawimy portrety fazowe kilku systemów: indeksu giełdowego DJIA, jak też kursu wymiany dolara amerykańskiego USD wobec marki niemieckiej DEM.



Jak widać, na wykresach są obecne pewne struktury, które nie przypominają zamkniętych orbit. Bliższe przyjrzenie się tym wykresom pozwala jednak zauważyć pewną regularność: rekonstruowane trajektorie w przestrzeni fazowej poruszają się zgodnie ze wskazówkami zegara po nieregularnym okręgu (elipsie lub innym konturze owalnym). Efekt ten jest najlepiej widoczny dla rekonstrukcji trajektorii DJIA. Aby pokazać go wyraźniej, postanowiliśmy posłużyć się notowaniami tygodniowymi i odszumiliśmy te dane za pomocą średniej kroczącej o rozmiarze okna 20. Widać tam wyraźne orbity zakreślane mniej więcej wokół punktu (0,0) na układzie współrzędnych. Jest to pewna przesłanka do założenia istnienia chaosu deterministycznego w tym układzie. Nie jest to

jednak ostateczny dowód, podobnie jak nieregularne kształty trajektorii na pozostałych portretach fazowych nie są zaprzeczeniem tezy o chaosie na rynkach kapitałowych. Świadczą one jedynie o obecności dużej ilości szumu w tych układach. Być może jednak układy te są mimo wszystko sterowane jakimś wewnętrznym deterministycznym i nieliniowym mechanizmem, który jest odpowiedzialny za tworzenie się regularnych trajektorii, jak w najbardziej wyrazistym do tej pory przypadku indeksu DJIA. Trzeba jeszcze podkreślić, że pozo-



stałe badane szeregi czasowe również mają trajektorie obracające się zgodnie ze wskazówkami zegara, jednak bądź nie są one domknięte (WIG), bądź obserwujemy na nich tylko jedno okrążenie (kurs wymiany USD-DEM), oraz skupione na niewielkim obszarze zakłębienie trajektorii. Przyczyną tych efektów jest prawdopodobnie mała wielkość szeregów czasowych, jakich użyliśmy (dla WIG-u jedynie około 10 lat z oczywistych względów, dla kursów USD-DEM z trzydziestu lat, ponieważ notowania z wcześniejszego okresu były poniekąd sztuczne ze względu na utrzymywanie się systemu z Bretton Woods, zakładającego stały z dokładnością do niewielkich odchyleń kurs walut wobec dolara amerykańskiego). Zatem konkluzją z próby rekonstrukcji trajektorii w przestrzeni fazowej badanych szeregów czasowych jest to, że zaobserwowano pewne cechy świadczące o możliwości istnienia chaosu deterministycznego na rynkach kapitałowych.

Podsumowanie

W pracy przedstawiliśmy podstawowe idee nieliniowej analizy szeregów czasowych oraz ich zastosowanie w badaniach ekonomicznych. W części teoretycznej opisaliśmy kilka technik analizy użytych następnie do pracy z danymi pochodzącymi z rynków walutowych i giełd papierów wartościowych. Zaprezentowaliśmy odtworzone trajektorie badanych systemów w przestrzeni fazowej, uzyskując ciekawe obrazy mające pewne znamiona regularności (obracanie się trajektorii zgodnie z ruchami wskazówek zegara). Oszacowaliśmy również maksymalny wykładnik Lyapunova metodą Kantza. Wyniki wykazały istnienie dodatniego wykładnika dla kursu wymiany USD-DEM, co jest bardzo silną przesłanką, popierającą tezę o istnieniu dynamiki nieliniowej w kursach wymiany walut. Dla porównania oszacowaliśmy również wykładnik Lyapunova dla surogatu szeregu czasowego USD-DEM otrzymując diametralnie różny obraz.

Literatura

1. P. CHEN, Empirical and theoretical evidence of economic chaos, *System Dynamics Review*, Vol. 4, str. 81 (1988).
2. A.M. FRASER, H.L. SWINNEY, Independent coordinates for strange atractors from mutual information, *Physical Review A*, Vol. 33, str. 1134 (1986).
3. R. HEGGER, H. KANTZ, TH. SCHREIBER, Practical implementation of nonlinear time series methods:The TISEAN package, *Chaos*, Vol. 9, str. 413 (1999).
4. H. KANTZ, A robust method to estimate the maximal Lyapunov exponent from a time series, *Physics Letters A*, Vol. 185, str. 77 (1994).
5. H. KANTZ, TH. SCHREIBER, *Nonlinear Time Series Analysis,* Cambridge University Press, Cambridge, 2000.
6. M.B. KENNEL, R. BROWN, H.D.I. ABANDEL, Determining embedding dimension for phase space reconstruction using a geometrical construction, *Physical Review A*, Vol. 45, str. 3403 (1992).
7. E.E. PETERS, *Teoria chaosu a rynki kapitałowe,* WIG-Press, Warszawa 1997.

8. TH. SCHREIBER, Interdisciplinary application of nonlinear time series methods, *Physics Reports*, Vol. 308, str. 1 (1999).
9. J. THEILER, S. EUBANK, A. LONGTIN, B. GALDRIKIAN, J.D. FARMER, Testing for nonlinearity in time series: The method of surrogate data, *Physica D*, Vol. 58, str. 77 (1992).

Methods of deteministic chaos in economical research

Abstract

An introduction to nonlinear analysis of economical time series is given. Methods developed for investigations of chaotic dynamics of physical systems are applied to various indexes and exchange rates. An attempt to distinguish between well-developed and emerging markets is undertaken. Phase-space reconstruction as well as estimation of the Lyapunov exponent are performed using available economical data.

Elżbieta Saganowska
Katedra Ekonometrii i Informatyki SGGW

Szacunek produktu krajowego brutto według 16 województw w 1998 roku[1]

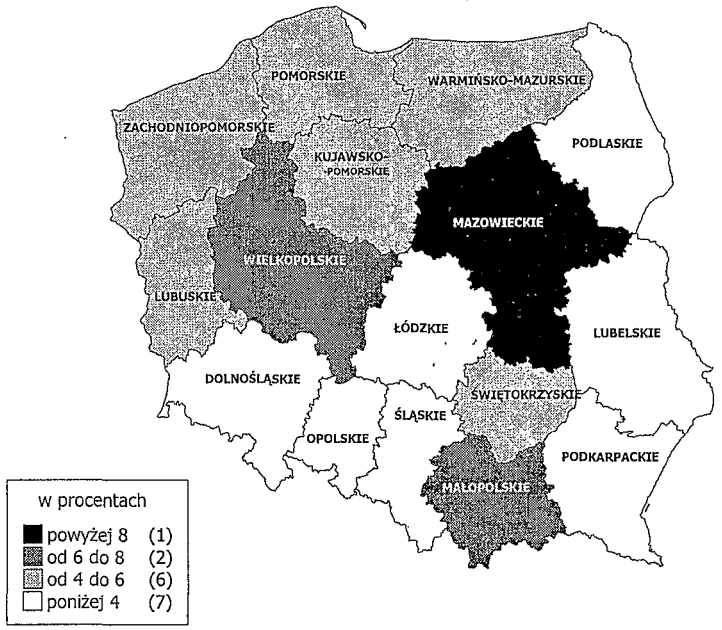
Niniejsze opracowanie zawiera wyniki badania dotyczącego szacunków regionalnego produktu brutto w 1998 roku dla 16 województw. Po raz pierwszy uwzględniono w szacunkach PKB szarą strefę gospodarki, bazując na skorygowanych o szacunki szarej strefy danych publikowanych przez GUS. Weryfikacja szacunków regionalnego PKB objęła lata 1995–1998.

Szacunki mają charakter eksperymentalny. Zostały przeprowadzone przy użyciu uproszczonej metodologii, której zastosowanie pozwala na szybsze uzyskanie wyników niż ma to miejsce w przypadku zastosowania metody pełnej. Najogólniej mówiąc, prace polegały na oszacowaniu, na podstawie szybko dostępnych danych statystycznych, poziomu produktu krajowego brutto w 1998 roku według 49 województw, a następnie opracowaniu (przy użyciu metod regresji liniowej) uśrednionych w czasie korekt dopasowujących te szacunki do obserwowanych wartości z lat 1995–1997. Szacunkowe dane dla 49 województw posłużyły do obliczenia wielkości tej kategorii w układzie 16 „nowych" regionów.

Szacunek realnego tempa wzrostu PKB w 1998 roku

Spośród 16 województw Polski zdecydowanie najwyższym tempem wzrostu gospodarczego (tempem wzrostu PKB w cenach stałych) wyróżnia się wysoko rozwinięte województwo mazowieckie (rys. 1). Różnica między nim a kolejnymi dwoma województwami o relatywnie wysokim tempie wzrostu gospodarczego – województwem wielkopolskim i województwem małopolskim – wynosi odpowiednio 2,4 i 3,8 punkty procentowe, podczas gdy różnice pomiędzy pozostałymi regionami wahają się od 0,05 do 1,8 punktu procentowego.

[1]Tekst referatu opracowany na podstawie [3].



Rysunek 1
Szacunek realnego tempa wzrostu PKB w 1998 roku

Silny wzrost gospodarczy charakteryzuje województwa położone w północnej i zachodniej części kraju (z wyjątkiem województwa dolnośląskiego) oraz województwo świętokrzyskie. Czołową pozycję wśród tych regionów zajmuje województwo warmińsko-mazurskie (5,6%).

Stosunkowo powolny wzrost charakteryzuje regiony tradycyjnie stanowiące obszary przemysłowe Polski (województwa: dolnośląskie – 0,9%, opolskie – 1,3%, śląskie – 1,7% i łódzkie – 2,0%). Słaby wzrost gospodarczy daje się również zauważyć w województwach podkarpackim (2,3%) i podlaskim (2,2%). Jedynym regionem, w którym według naszych szacunków nastąpił spadek (około 1%) realnego tempa wzrostu gospodarczego w 1998 roku jest województwo lubelskie. Relatywnie niski wzrost gospodarczy w województwach usytuowanych na „wschodniej ścianie" Polski można tłumaczyć osłabieniem współpracy przygranicznej.

Wobec spowolnienia tempa wzrostu gospodarczego Polski w 1998 r. w stosunku do lat 1995–1997 o 1–2 punkty procentowe, analogiczne tendencje wystąpiły również na poziomie regionalnym. Wyjątek stanowiły dwa woje-

wództwa: mazowieckie i kujawsko-pomorskie, w których szacowane dane odnośnie realnego tempa wzrostu gospodarczego wykazywały wyższy (o 1–2 punkty procentowe) wzrost w 1998 r. w porównaniu z rokiem poprzednim.

Szacunek PKB na 1 mieszkańca w 1998 roku

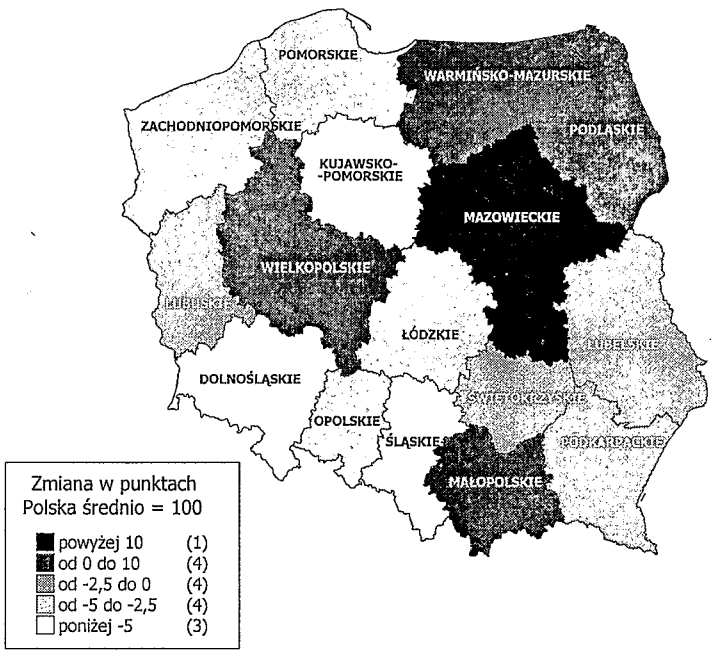
Analiza poziomu rozwoju gospodarczego województw na podstawie szacunkowych danych o wartości PKB na 1 mieszkańca wskazuje, iż jedynie w trzech województwach Polski: mazowieckim, wielkopolskim i śląskim, wartość produktu krajowego brutto w przeliczeniu na 1 mieszkańca kształtowała się w 1998 roku na poziomie wyższym od średniej krajowej i wynosiła w relacji do średniej krajowej odpowiednio 151,5%, 111,2% oraz 105,3%. Pozostałe regiony lokowały się w trzech grupach, o wartościach wskaźnika zawartych w granicach:

Tabela 1
Szacunek wartości produktu krajowego brutto na 1 mieszkańca w latach 1995–1998

Województwa	Wartość per capita w zł				Średnia krajowa = 100			
	1995	1996	1997	1998	1995	1996	1997	1998
Polska	7934	9976	12141	14210	100,0	100,0	100,0	100,0
Dolnośląskie	8101	9908	12189	13747	102,1	99,3	100,4	96,7
Kujawsko-pomorskie	7420	9029	10632	12481	93,5	90,5	87,6	87,8
Lubelskie	5691	7369	8938	9906	71,7	73,9	73,6	69,7
Lubuskie	6988	8714	10466	12268	88,1	87,4	86,2	86,3
Łódzkie	7551	9379	11244	12873	95,2	94,0	92,6	90,6
Małopolskie	7624	9485	11746	13918	96,1	95,1	96,7	97,9
Mazowieckie	10710	14139	17529	21528	135,0	141,7	144,4	151,5
Opolskie	6873	8489	10325	11708	86,6	85,1	85,0	82,4
Podkarpackie	6048	7660	9254	10547	76,2	76,8	76,2	74,2
Podlaskie	5931	7619	9435	10767	74,8	76,4	77,7	75,8
Pomorskie	8268	10159	12145	14200	104,2	101,8	100,0	99,9
Śląskie	8903	10990	13151	14961	112,2	110,2	108,3	105,3
Świętokrzyskie	5979	7474	8877	10334	75,4	74,9	73,1	72,7
Warmińsko-mazurskie	6132	7757	9695	11424	77,3	77,8	79,9	80,4
Wielkopolskie	8543	10721	13177	15807	107,7	107,5	108,5	111,2
Zachodniopomorskie	7844	9708	11690	13634	98,9	97,3	96,3	95,9

- od 90 do 100% – województwa: pomorskie, małopolskie, dolnośląskie, zachodniopomorskie, łódzkie,
- od 80 do 90% – województwa: kujawsko-pomorskie, lubuskie, opolskie, warmińsko-mazurskie,
- od 70 do 80% – województwa: podlaskie, podkarpackie, świętokrzyskie i lubelskie.

Analiza stopnia rozwoju województw określona przez zmianę relacji poziomu PKB na 1 mieszkańca w regionach do średniego poziomu PKB na 1 mieszkańca w Polsce wskazuje, iż grupa województw wykazujących w latach 1995–1998 awans gospodarczy w stosunku do reszty kraju (wzrost nominalnego



Rysunek 2
Relatywna zmian poziomu PKB na 1 mieszkańca w latach 1995–1998

PKB szybszy od średniego) zmniejszyła się w porównaniu z okresem 1995–1997. W grupie województw „awansujących" pozostały: mazowieckie, wielkopolskie, warmińsko-mazurskie, małopolskie i podlaskie, natomiast wcześniej wchodzące do tej grupy województwo lubelskie „przeszło" do kategorii województw o gorszej sytuacji w stosunku do średniej krajowej.

Zauważyć należy także, iż w latach 1995–1998 dystans między poziomem PKB na 1 mieszkańca w pozostałych regionach nie zmienił się w porównaniu z okresem 1995–1997, a w przypadku niektórych spośród nich (łódzkie, opolskie, śląskie, dolnośląskie) zwiększył się znacząco (od 2,0 do 3,7 punktów procentowych) na ich niekorzyść.

Obserwowany w Polsce na przestrzeni lat 1995–1997 silny wzrost gospodarczy, w granicach 6–7% rocznie, w roku 1998 uległ osłabieniu. Analogicznie, w większości regionów kraju zmniejszyło się tempo rozwoju gospodarczego (w województwie lubelskim nastąpił nawet jego spadek). Tendencji tej oparły się jedynie dwa województwa – mazowieckie oraz kujawsko-pomorskie, w których dynamika produktu krajowego brutto w przeliczeniu na 1 mieszkańca w cenach stałych w 1998 roku była wyższa w porównaniu z wartościami tego wskaźnika z 1997 roku odpowiednio o 1,2 i 1,8 punktu procentowego.

Tabela 2
Dynamika PKB w przeliczeniu na 1 mieszkańca w latach 1996–1998 (ceny stałe)

Województwa	1996	1997	1998
	Rok poprzedni = 100		
Dolnośląskie	103,0	107,9	101,0
Kujawsko-pomorskie	102,5	103,3	105,1
Lubelskie	109,1	106,4	99,2
Lubuskie	105,0	105,3	104,9
Łódzkie	104,6	105,1	102,5
Małopolskie	104,8	108,6	106,1
Mazowieckie	111,2	108,7	110,0
Opolskie	104,0	106,7	101,5
Podkarpackie	106,7	106,0	102,0
Podlaskie	108,2	108,6	102,2
Pomorskie	103,5	104,9	104,7
Śląskie	104,0	104,9	101,8
Świętokrzyskie	105,3	104,2	104,2
Warmińsko-mazurskie	106,6	109,6	105,5
Wielkopolskie	105,7	107,8	107,4
Zachodniopomorskie	104,3	105,6	104,4

Największe osłabienie tempa rozwoju gospodarczego (spowolnienie wzrostu PKB na 1 mieszkańca od ok. 4 do ok. 7 punktów procentowych) wystąpiło w czterech województwach: podkarpackim, warmińsko-mazurskim, opolskim i dolnośląskim.

Sytuacja na rynku pracy

W latach 1996–1998 nastąpiła nieznaczna poprawa na rynku pracy w Polsce – stopa rejestrowanego bezrobocia spadła o 2–3 punkty procentowe. Wśród 16 województw aż w piętnastu stopa bezrobocia zmniejszyła się, przy czym spadek ten był nierównomierny i wahał się od 1,8 (województwo śląskie) do 5,3 punktów (województwo lubelskie). Jedynie w województwie lubuskim nastąpił wzrost tego wskaźnika (0,6 punktu).

Pozytywna tendencja na rynku pracy najwyraźniej widoczna jest w województwach centralnej i północno-wschodniej Polski (spadek stopy bezrobocia w granicach 4–5,3 punktów) i dotyczy przede wszystkim regionów charakteryzujących się wysokim poziomem bezrobocia w 1996 r. Przykładowo, województwo warmińsko-mazurskie, w którym wystąpił 4-punktowy spadek stopy bezrobocia, miało w analizowanym okresie najwyższy poziom wskaźnika w kraju. Podobną charakterystykę odnieść można do województw kujawsko-pomorskiego i łódzkiego.

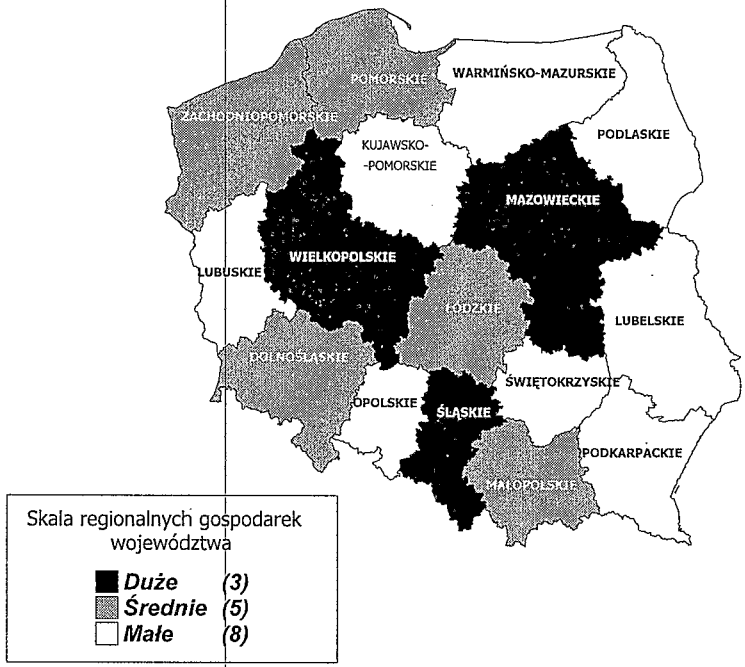
Charakterystyka 16 województw ze względu na wybrane kategorie społeczno-ekonomiczne

Celem reformy administracyjnej Polski było stworzenie możliwie silnych regionów, których samorządy, dzięki przekazanym z centrum kompetencjom, mogłyby prowadzić politykę zmierzającą do racjonalnego wykorzystywania wypracowanych i otrzymanych z zewnątrz środków finansowych oraz skutecznego rozwiązywania wielu społecznych i gospodarczych problemów lokalnych.

Podjęliśmy, opierając się na wynikach przeprowadzonego badania, próbę oceny potencjału gospodarczego nowych 16 województw w 1998 roku. Siłę gospodarczą i poziom rozwoju gospodarczego regionu najlepiej odzwierciedlają wielkości:

- produktu krajowego brutto tworzonego w danym województwie,
- produktu krajowego brutto tworzonego w danym województwie w przeliczeniu na 1 mieszkańca tego województwa.

Wyniki badania sugerują ukształtowanie się w nowym podziale administracyjnym kraju trzech grup województw, które ze względu na skalę regionalnych gospodarek zostały zdefiniowane jako „duże", „średnie" oraz „małe". Pod względem liczby ludności grupy te nie różnią się znacząco; każda z nich absorbuje ok. jednej trzeciej ogółu ludności Polski.



Rysunek 3
Grupy województw według skali regionalnych gospodarek

Pierwszą grupę stanowią województwa: mazowieckie, wielkopolskie i śląskie, zajmujące ok. jednej czwartej obszaru kraju. Są to regiony, w których wytwarza się około 43% PKB Polski. Województwa te cechuje również najwyższy poziom PKB w przeliczeniu na 1 mieszkańca. W relacji do średniej krajowej był on przeciętnie wyższy o ok. 23% i wahał się od 5,3% (województwo śląskie) do 51,5% (województwo mazowieckie). Stopa rejestrowanego bezrobocia w województwach tworzących tę grupę nie przekracza 8% i należy do najniższych w kraju.

Drugą grupę województw – „średnia" skala regionalnych gospodarek – stanowią województwa: dolnośląskie, łódzkie, małopolskie, pomorskie i zachodniopomorskie, zajmujące ok. 30% powierzchni Polski. Wytwarza się w nich 31,8% PKB kraju, przy czym liderami w grupie są województwa małopolskie (8,2%) i dolnośląskie (7,6%). Wielkość PKB w przeliczeniu na 1 mieszkańca kształtuje się w tej grupie województw nieznacznie poniżej pozio-

mu średniego dla kraju (od 96 do 99,9%), z wyjątkiem łódzkiego, w którym wielkość PKB per capita wynosi 90,6%. Stopa rejestrowanego bezrobocia jest w tej grupie województw wyższa niż w grupie pierwszej przeciętnie o 3–4 punkty procentowe. Tylko w województwie małopolskim wielkość tego wskaźnika jest porównywalna ze stopą bezrobocia notowaną w województwach pierwszej grupy.

Tabela 3
Grupy województw według skali regionalnych gospodarek

Województwa	Udział pow. województwa w pow. Polski	Liczba ludności w mln	Udziały tworzeniu PKB w %	Poziom PKB per capita, średnia krajowa = 100	Stopa bezrobocia
Duże					
Mazowieckie					
Śląskie	24,9	3,3–5,1	9,6–20	105–152	7,3–8,0
Wielkopolskie					
Średnie					
Dolnośląskie					
Łódzkie					
Małopolskie	30,2	2,0–3,2[a)]	5,6–8,2[a)]	91–100	11,0–13,9[b)]
Pomorskie					
Zachodniopomorskie					
Małe					
Kujawsko-pomorskie					
Lubelskie					
Lubuskie					
Opolskie	44,9	1,0–2,2	poniżej 4,8	70–90	10,3–19,7
Podkarpackie					
Podlaskie					
Świętokrzyskie					
Warmińsko-mazurskie					

[a)] z wyjątkiem zachodniopomorskiego, gdzie udział w tworzeniu PKB wynosił 4,3%, a liczba ludności 1,7 mln.
[b)] z wyjątkiem małopolskiego, gdzie stopa bezrobocia wynosi 7,6.

Łącznie w tych dwóch grupach, skupiających połowę regionów kraju, wytwarzane jest trzy czwarte produktu krajowego brutto.

Trzecią grupę województw, relatywnie najsłabszych gospodarczo, tworzy osiem pozostałych regionów, zajmujących ok. 45% powierzchni Polski. Są to

województwa: kujawsko-pomorskie, lubelskie, lubuskie, opolskie, podkarpackie, podlaskie, świętokrzyskie i warmińsko-mazurskie. Wytwarzany w nich PKB stanowi ok. 25% wartości krajowej. W stosunku do dwóch poprzednich grupę tę charakteryzuje relatywnie niski poziom rozwoju gospodarczego. Poziom PKB w przeliczeniu na 1 mieszkańca był niższy o 10–30% w relacji do średniej krajowej, przy czym aż w czterech województwach tej grupy wynosił mniej niż 76% tej wartości. W województwach tych stopa bezrobocia jest porównywalna do wielkości wskaźnika wykazywanego w województwach zaliczonych do średniej grupy. Jedynie region warmińsko-mazurski wyróżnia się bardzo wysokim, nie mającym odpowiednika w kraju, bezrobociem (19,7%).

Aneks: Metodologia szacunku

Podstawą do opracowania metody pozwalającej na oszacowanie poziomu produktu krajowego brutto w układzie 16 województw w 1998 roku były wyniki badania kształtowania się tej kategorii według 49 województw w latach 1995–1997[2] opublikowane przez GUS oraz szacunki ZBS-E dla 1998 roku.

Badanie przeprowadzono w dwóch etapach:

- w pierwszym – na podstawie uproszczonej metodologii oszacowano poziom produktu krajowego brutto, wartości dodanej brutto ogółem oraz wartości dodanej brutto w poszczególnych sektorach ekonomicznych w cenach bieżących w układzie 49 województw w 1998 roku;
- w drugim – uzyskane szacunki przedstawiono w układzie 16 województw[3].

Wstępne[4] wojewódzkie szacunki wartości dodanej w rolnictwie, przemyśle, budownictwie, usługach rynkowych i usługach nierynkowych zostały skorygowane przy użyciu metody funkcji regresji liniowej. Podstawę tej korekty stanowiły dane dotyczące poziomu wartości dodanej brutto w poszczególnych sektorach ekonomicznych w układzie 49 województw w latach 1995–1997. Do oszacowania przyjęto funkcję regresji następującej postaci:

---

[2]Por. „Produkt krajowy brutto według województw za lata 1995–1997", GUS i US w Katowicach, październik 1999.

[3]Por. W. Orłowski, E. Saganowska, L. Zienkowski, „Szacunek produktu krajowego brutto według 16 województw za 1996 i 1997 rok (Metoda uproszczona)" Studia i Prace, z Prac ZBS-E GUS i PAN, zeszyt nr 262, Warszawa 1998.

[4]Por W. Orłowski, E. Saganowska, T. Śmiłowska, „Eksperymentalny szacunek wzrostu produktu krajowego brutto według województw w 1996 r", Studia i Prace, z Prac ZBS-E GUS i PAN, zeszyt nr 252, Warszawa 1997. [w:] „Wybrane problemy regionalnego zróżnicowania Polski".

$$y_{ij}^{t} = \alpha_i + \gamma_{ij} \times u_{ij}$$

gdzie: y_{ij}^{t} – względny procentowy błąd szacunku zdefiniowany jako

$$y_{ij}^{t} = \frac{y_{ij}^{DANE(t)} - y_{ij}^{SZAC(t)}}{y_{ij}^{DANE(t)}} \times 100$$

$y_{ij}^{DANE(t)}$ – wartość dodana brutto w i-tym sektorze ekonomicznym, j-tym województwie – dane ZBS-E i US w Katowicach,

$y_{ij}^{SZAC(t)}$ – wstępne szacunki dla i-tego sektora ekonomicznego, j-tego województwa,

α_i – stały parametr funkcji regresji dla i-tego sektora ekonomicznego,

γ_{ij} – parametr stojący przy zmiennej 0-1, i-ty sektor ekonomiczny, j-te województwo,

u_{ij} – zmienna zero-jedynkowa przypisana do i-tego sektora ekonomicznego, w j-tym województwie,

i = rolnictwo, przemysł, budownictwo, usługi rynkowe i usługi nierynkowe (sektory ekonomiczne),

j = 1 ... 49,

t = 1995, 1996, 1997.

Oszacowane funkcje regresji oraz wstępne szacunki wojewódzkich wartości dodanych dla poszczególnych sektorów ekonomicznych posłużyły do opracowania prognoz wartości dodanej w rolnictwie, przemyśle, budownictwie, usługach rynkowych i usługach nierynkowych na rok 1998.

Wartości dodane dla Polski dla analizowanych sektorów różniły się nieznacznie od wartości rocznikowych. Różnice między oszacowaną wartością dodaną poszczególnych sektorów dla Polski a analogiczną wartością podawaną przez GUS rozszacowano przy użyciu wskaźników określających udział wojewódzkiej teoretycznej wartości dodanej w teoretycznej wartości dodanej dla kraju.

Literatura

1. W. ORŁOWSKI, E. SAGANOWSKA, L. ZIENKOWSKI, Szacunek produktu krajowego brutto według 16 województw za 1996 i 1997 rok (Metoda uproszczona), *Studia i Prace, z Prac ZBS-E GUS i PAN*, zeszyt nr 262, Warszawa 1998.

2. W. ORŁOWSKI, E. SAGANOWSKA, T. ŚMIŁOWSKA, Eksperymentalny szacunek wzrostu produktu krajowego brutto według województw w 1996 r., *Studia i Prace, z Prac ZBS-E GUS i PAN*, zeszyt nr 252, Warszawa 1997. [w:] *Wybrane problemy regionalnego zróżnicowania Polski.*
3. E. SAGANOWSKA, GÓRALCZYK-MODZELEWSKA, M. WOJCIECHOWSKA, Zróżnicowanie produktu krajowego brutto według województw w 1998 roku na podstawie oszacowań metodą uproszczoną, www.GUS.gov.stad.pl.
4. *Biuletyn Statystyczny*, nr 12/1999, GUS, Warszawa 1999.
5. *Rocznik Statystyczny 1999*, GUS, Warszawa 1999.
6. *Mały Rocznik Statystyczny 1998*, GUS, Warszawa 1998.
7. *Mały Rocznik Statystyczny 1999*, GUS, Warszawa 1999.
8. Produkt krajowy brutto według województw za lata 1995–1997, GUS i Urząd Statystyczny w Katowicach, Katowice 1999.
9. *Rocznik Statystyczny Województw 1999*, GUS, Warszawa 1999.
10. Dane uzyskane bezpośrednio w GUS w Departamencie Pracy.
11. Dane z Banku Danych Lokalnych GUS.

Estimates of GDP in new voivodships in 1998

Abstract

Paper presents estimates of regional GDP in new 16 voivodships for the years 1996–1998. Estimates were constructed using the newly elaborated „simplified" methodology, which was based on the available, published regional data. The advantage of the simplified approach is that it produces regional GDP estimates well before the „official" data are available.

Results show that more than 40% of Poland's GDP in 1998 originated from three the most developed voivodships: mazowieckie, wielkopolskie and salskie. The second group of voivodships, with the average GDP per capita levels produced more than 30% of Poland's GDP. Other 8 voivodships produced ¼ of Poland's GDP.

Maria Parlińska
Katedra Ekonometrii i Informatyki SGGW

Statystyczna kontrola jakości

W gospodarce rynkowej głównym kryterium leżącym u podstaw wyborów wariantów produkcji jest wynik ekonomiczny. Znaczenie ekonomiczne problematyki jakości z punktu widzenia przedsiębiorstw kształtowane jest na podstawie sum strat i kosztów związanych z jakością produkcji[1]. Aby zatem osiągnąć jak największe zyski, należy dążyć do poprawy jakości wytwarzanego produktu. Dzięki wysokiej jakości produktu zmniejszamy lub wręcz eliminujemy częstotliwość występowania sytuacji, w których ponosimy koszty na usuwanie wad technicznych, naprawy i remonty, straty z tytułu reklamacji itp.

Jakością danego produktu nazywa się zbiór cech przysługujących (przynależnych) temu produktowi. Ustalenie jakości dowolnego produktu polega na odkryciu i sformułowaniu zbioru cech odniesionych do danego produktu[2] w trakcie procesu nazywanego „kontrolą jakości".

W zależności od rozmiarów jednocześnie sprawdzanej produkcji rozróżnia się **kontrolę całkowitą** i **procentową**.

Kontrolę całkowitą (zupełną albo stuprocentową), polegającą na sprawdzeniu wszystkich jednostek zgłoszonej partii, stosuje się w następujących przypadkach:

a) gdy sprzęt produkcyjny i właściwości procesu technologicznego nie mogą zapewnić jednorodności parametrów jakościowych produktów (procesy nieustabilizowane o dużym rozrzucie tych parametrów);
b) po operacjach mających decydujące znaczenie dla jakości dalszej obróbki lub montażu;
c) przy sprawdzaniu szczególnie ważnych i kosztownych produktów lub których wadliwe wykonanie zagraża bezpieczeństwu życia użytkownika;
d) dla rozsegregowania na grupy selekcyjne przed montażem albo dla wyeliminowania wadliwych elementów z partii nie przyjętych przy kontroli procentowej.

Kontrola procentowa, polegająca na pobraniu pewnej ściśle określonej liczby elementów z partii (wyznaczonej procentem od ilości zgłoszonej),

---

[1]Lesław Wasilewski – *Metody kontroli jakości w przedsiębiorstwach przemysłowych.*

[2]Adam Hamrol, Władysław Mantura – *Zarządzanie jakością, teoria i praktyka.*

jest zwykle stosowana w następujących przypadkach:

a) gdy sprzęt produkcyjny lub przebieg procesu technologicznego zapewnia dużą jednorodność parametrów jakościowych (proces ustabilizowany), których zmienność przebiega w wąskich granicach (np. obróbka na automatach);
b) po operacjach, które nie dają ostatecznych parametrów jakościowych i nie wykazują dużych ilości braków (np. obróbka wstępna);
c) przy odbiorze dużych partii jednakowych elementów, a zwłaszcza mniej ważnych.

Kontrola procentowa powinna być dokonywana zgodnie z zasadami **metod statystycznej kontroli jakości.**

Cel statystycznej kontroli jakości

Nowoczesna kontrola jakości nie może ograniczać się do badania zgodności wykonywanych produktów z dokumentacją konstrukcyjną wraz z warunkami technicznymi i eliminacji braków w celu niedopuszczenia do użytkowników produktów o niewystarczającej jakości i nieprzydatnych w eksploatacji.

Nowoczesna KJ powinna poza tymi zadaniami aktywnie uczestniczyć w wykrywaniu braków i przyczyn ich powstawania w trakcie trwania procesu produkcyjnego, powinna sygnalizować możliwość pojawienia się braków i usunąć przyczynę, która może je spowodować nim się pojawią. Nowoczesna KJ powinna również stosować metody najbardziej ekonomiczne, tzn. koszty kontroli powinny być mniejsze niż koszt braków, które powstałyby, gdyby kontroli nie było (od zasady tej można odstępować tylko w wyjątkowych przypadkach, np. gdy braki mogą zagrażać życiu ludzkiemu, gdy chodzi nam o względy wychowawcze itp.).

Przy stosowaniu tradycyjnej kontroli stuprocentowej spełnienie tych wymagań napotyka na szereg trudności, szczególnie przy produkcji wielkoseryjnej i masowej, coraz bardziej dominującej obecnie w przemyśle.

Zachodzi wtedy konieczność wielokrotnego badania wielkiej liczby przedmiotów, co wymaga rozbudowy aparatu kontroli, organizowania dużej liczby stanowisk kontrolnych. Poza tym kontrola stuprocentowa nie jest w stanie spełnić w pełni wymagań stawianych nowoczesnej KJ, ponieważ:

a) nie daje całkowitej pewności bezbłędnej kontroli;
b) nie daje możliwości zapobiegania powstawaniu braków, gdyż wykonywana jest bądź w odniesieniu do całej partii kontrolowanych produktów po zakończeniu określonej operacji przy kontroli międzyoperacyjnej, bądź do

partii gotowych produktów przy kontroli ostatecznej i odbiorczej;

c) jest bardzo pracochłonna, szczególnie w przypadkach konieczności znacznego rozszerzenia kontroli międzyoperacyjnej;

d) jest czasem niemożliwa (np. przy próbach niszczących).

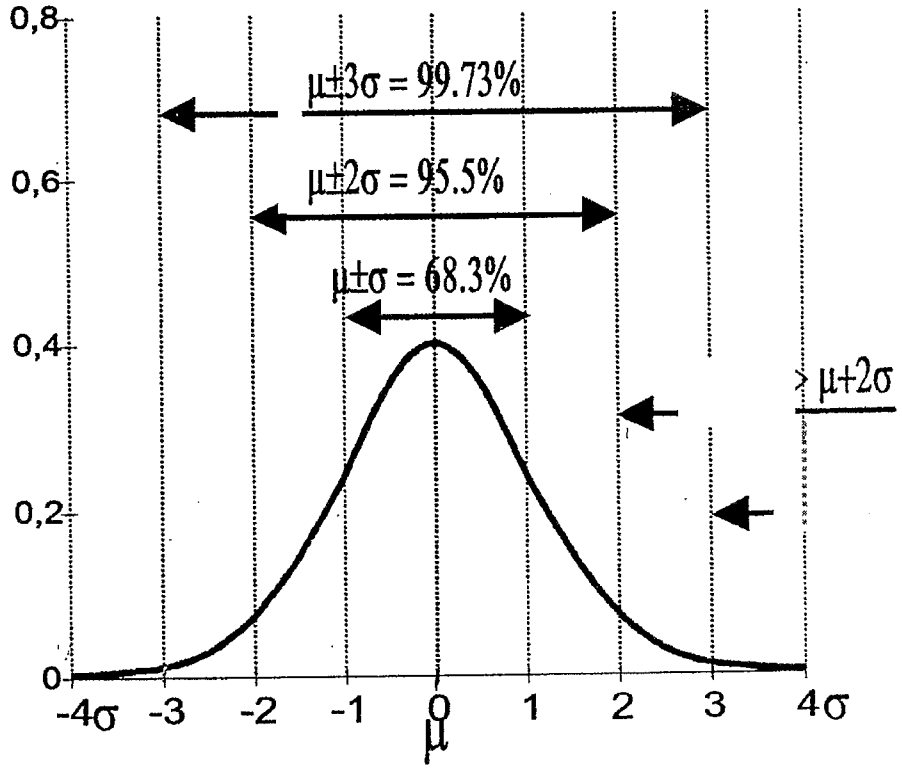
Wymagania nowoczesnej kontroli spełnia **statystyczna kontrola jakości** (skrót **SKJ**), opierająca się na *statystyce matematycznej i rachunku prawdopodobieństwa*. Istota jej polega na poddawaniu badaniu nie całej partii przedmiotów, lecz pewnej części tej partii (tzw. próbki) i orzekaniu na tej podstawie o jakości całej partii. Oczywiście wnioski wyciągnięte na podstawie badania próbki mogą się różnić od stanu faktycznego, jednak wielkość ryzyka podjęcia mylnej decyzji jest dokładnie znana i co więcej – można, dobierając odpowiednio warunki kontroli, świadomie ustalić wielkość tego ryzyka, a to właśnie dzięki zastosowaniu zasad rachunku prawdopodobieństwa. Moment ten wskazuje na wyższość SKJ nad kontrolą stuprocentową, przy której nie jest znana wielkość ryzyka podjęcia mylnej decyzji. O wyższości SKJ nad kontrolą stuprocentową świadczy również jej opłacalność, ze względu na zasadnicze zmniejszenie pracochłonności badania przy kontroli odbiorczej i wyraźnie profilaktyczny charakter przy kontroli bieżącej, ponieważ badanie odbywa się w czasie trwania procesu produkcyjnego, a wynik tego badania jest podstawą do natychmiastowej decyzji zatrzymania procesu, gdy przestał on być ustabilizowany, tzn. uległ rozregulowaniu lub też jest na granicy rozregulowania się.

Nazwa **statystyczna kontrola jakości** pochodzi stąd, że przy badaniu zbioru wyrobów, zespołów czy części posługuje się do ich oceny metodami statystyki matematycznej. Rachunek prawdopodobieństwa jest podstawą do określenia wielkości ryzyka podjęcia mylnej decyzji, ze względu na podejmowanie jej na podstawie badania nie całego zbioru, a tylko jego części.

Cel badań normalności rozkładu

Wiele metod kontroli zarówno odbiorczej, jak i bieżącej opartych jest na założeniu, że rozkłady badanych cech są rozkładami normalnymi. Dlatego też przed przystąpieniem do stosowania metod statystycznych wstępną czynnością jest sprawdzenie, czy cecha (zmienna), która ma być poddana badaniom, podlega rozkładowi normalnemu.

Ponadto, obowiązkiem dostawcy, jeśli odbiorca będzie tego wymagał, jest zapewnienie i udowodnienie, że cechy dostarczanych przez niego produktów tworzą w partiach rozkłady normalne.



Na rysunku powyżej przedstawiono oczekiwane procentowe ilości wystąpień w przedziałach $\mu \pm 1\sigma$, $\mu \pm 2\sigma$, $\mu \pm 3\sigma$.

Przedmiot i zakres prowadzenia karty

Karta kontrolna „$\overline{X} - R$" wartości średniej $\overline{X}$ i rozstępu R w próbce służy do kontroli zmienności cech, które mogą być oceniane liczbowo. Do takich cech należą: wymiar, masa, gęstość, wytrzymałość, temperatura itp.

W praktyce zachodzi często konieczność kontrolowania kilku cech (na przykład kilku wymiarów). Należy wówczas prowadzić osobną kartę kontrolną dla każdej badanej cechy (zmiennej).

Teoria karty kontrolnej „$\overline{X} - R$" oparta jest na założeniu, że badana cecha ma rozkład normalny. W przypadkach kiedy stwierdzimy, że rozkład badanej zmiennej nie jest normalny, nie można prowadzić karty „$\overline{X} - R$" w sposób opisany w niniejszym opracowaniu.

Cel prowadzenia karty kontrolnej

Celem prowadzenia karty kontrolnej jest:

- kierowanie procesem technologicznym (produkcyjnym) na podstawie statystycznej analizy danych zebranych w czasie wytwarzania produktu. Informacje te pozwalają uchwycić rozregulowanie procesu w takim momencie, gdy nie prowadzi to jeszcze do powstania braków i umożliwiają podjęcie decyzji o sposobie regulacji procesu;
- ocena jakości produktu na podstawie analizy informacji zebranych w trakcie procesu wytwarzania. Jeżeli karta wskazuje, że proces nie jest uregulowany statystycznie, a produkt nie odpowiada stawianym wymaganiom, konieczne staje się przesortowanie, poprawienie lub ewentualne wybrakowanie produktów.

Budowa karty kontrolnej

Typowa karta kontrolna „$\overline{X} - R$" składa się z następujących części:

- części nagłówkowej (identyfikacyjnej) zawierającej dane dotyczące zakładu i wydziału produkcyjnego, urządzenia, detalu badanego, wykonywanej operacji, badanej cechy, osoby prowadzącej badania;
- części kalkulacyjnej, na której notuje się godziny pobierania próbek oraz wyniki mierzenia badanej cechy (zmiennej) X wszystkich sztuk wchodzących w skład próbki. Kolejne rubryki zawierają sumę $\sum X$ oraz rozstęp R w próbce;
- części graficznej, zawierającej tory dla dwóch parametrów statystycznych $\overline{X}$ i R. Na torach tych zaznaczone są punkty odpowiadające wyliczonym wcześniej wielkościom. Dodatkowo na wykresach naniesione są linie kontrolne określające graniczne wartości zmienności parametrów $\overline{X}$ i R.
- części pomocniczej, zawierającej uwagi o przebiegu procesu produkcyjnego (zakłóceniach i czynnościach regulacyjnych procesu), wymianie i ostrzeniu narzędzi, wynikach kontroli cechy nie będącej celem prowadzenia karty itp.

Wybór cech do badań

Przy wyborze cech wyrobu, które chcemy badać przy pomocy karty „$\overline{X} - R$", możemy skorzystać z podanych niżej wskazówek:

- wysiłki badawcze należy skupić na tych cechach (właściwościach), które najskuteczniej mogą wpłynąć na poprawę procesu i jakości wyrobu;
- należy brać pod uwagę potrzeby klienta. Niektóre właściwości wyrobu mogą być dla klienta szczególnie istotne i ważne. Czasem takie cechy są specjalnie zaznaczane w dokumentacji konstrukcyjnej (np. jako wymiary krytyczne);
- należy badać te właściwości, które w procesie produkcyjnym są powodem powstawania największych strat, braków, wiążą się z koniecznością napraw itp.

Ustalanie liczności próbki

Ważną czynnością przy zakładaniu karty kontrolnej jest ustalenie liczebności pobieranych próbek. Ma to decydujący wpływ na skuteczność i sprawność prowadzenia badań.

Liczebność próbki dla karty „$\overline{X} - R$" przyjmuje się w granicach od 2 do 10 sztuk. Próbkę o małej liczebności stosuje się, gdy badanie jest kosztowne, uciążliwe lub prowadzone jest metodą niszczącą. Na ogół przyjmuje się wielkość próbki w granicach od 4 do 6 sztuk. W przypadkach, gdy zależy nam na stwierdzeniu małych zmian w procesie produkcyjnym, możemy pobierać próbki zawierające 7 do 10 sztuk. Próbek większych nie stosuje się, gdyż powoduje to błędy przy ocenie rozstępu.

Należy zwrócić uwagę na wymóg, by pobierane próbki reprezentowały tylko jedno narzędzie itp. (tzw. pojedynczy strumień procesu).

Ustalanie odstępu miedzy próbkami

Odstęp między kolejnymi próbkami powinien być ustalony doświadczalnie. Postępowanie polega na kilkakrotnym prowadzeniu obserwacji procesu produkcyjnego przez pewien czas. Ma to na celu wykrycie momentów, w których następuje rozregulowanie procesu. Takie rozregulowania mogą być spowodowane na przykład tępieniem się narzędzia, nagrzewaniem, zmianą partii materiału itp.

Średni okres między rozregulowaniami – pomniejszony o pewien zapas bezpieczeństwa – przyjmujemy jako odstęp między próbkami. Jeżeli ten sam kontroler obsługuje kilka stanowisk roboczych, to wówczas ustalamy dla nich tę samą częstotliwość pobierania próbek.

W przypadku, gdy dla badanego procesu znamy jego zdolność jakościową, to możemy wówczas ustalać odstęp między próbkami przyjmując za kryterium

wskaźniki zdolności c_p i c_{pk} wg tabeli 1 (wg normy PN-ISO 3534-2 p. 3.2.6, wskaźnik zdolności c_p ma oznaczenie międzynarodowe CPI).

Należy brać również pod uwagę to, by zbyt częsta kontrola nie zwiększała bez potrzeby kosztów jej prowadzenia.

Ustalanie liczby próbek

Z punktu widzenia procesu produkcyjnego liczba pobieranych do badań próbek musi być wystarczająca, by zapewnić możliwość ujawnienia się głównych źródeł zmienności badanej cechy.

Jeżeli proces jest ustabilizowany, to dobrym sprawdzeniem stabilności jest zazwyczaj 25 lub więcej próbek, które zawierają 100 lub więcej indywidualnych pomiarów.

W pewnych przypadkach mogą być dostępne już istniejące dane, co umożliwia przyspieszenie pierwszej fazy badań procesu. Można je stosować tylko wówczas, gdy są to wyniki stosunkowo „świeże" i gdy kryteria wybierania próbek są dokładnie sprecyzowane.

Zakładanie i prowadzenie karty $\overline{X}-R$

Badania wstępne

Prowadzenie karty kontrolnej $\overline{X}-R$ jest możliwe w sytuacji, gdy zmienność badanej cechy (właściwości) X ma rozkład normalny. Dlatego przed rozpoczęciem badań pobieramy z bieżącej produkcji dużą próbkę i sprawdzamy normalność rozkładu. Do próbki należy pobierać kolejno sztuki z okresu, w którym można przyjąć niezmienność produkcji.

Następnie w ustalonych odstępach czasu pobieramy 20 do 30 próbek o określonej liczebności „n" (patrz: Ustalanie liczności próbki, Ustalanie odstępu miedzy próbkami). Dla każdej sztuki wyznaczamy wartość liczbową badanej cechy (np. średnicę itp.):

$$X_1, X_2, X_3, \ldots\ldots, X_n$$

Wartości te zapisujemy w części kalkulacyjnej arkusza karty. Dla każdej próbki o numerze „i" obliczamy średnią X_i ze wzoru:

$$\overline{X}_i = \frac{X_1 + X_2 + X_3 + ... + X_n}{n}$$

oraz rozstęp R_i ze wzoru:

$$R_i = X_{\max} - X_{\min}$$

gdzie $X_{\max}$ i $X_{\min}$ oznaczają największą i najmniejszą wartość liczbową cechy w próbce o numerze „*i*". Wartości $\overline{X}_i$ i R_i odnotowujemy w karcie.

Następnym krokiem postępowania jest obliczenie średniej ogólnej $\overline{\overline{X}}$ ze średnich z wszystkich próbek, posługując się wzorem:

$$\overline{\overline{X}} = \frac{\overline{X}_1 + \overline{X}_2 + \overline{X}_3 + ... + \overline{X}_n}{n}$$

oraz średniego rozstępu $\overline{R}$ ze wzoru:

$$\overline{R} = \frac{R_1 + R_2 + R_3 + ... + R_k}{k}$$

gdzie: k – jest liczbą próbek, R_1 i $\overline{X}_1$ to rozstęp i średnia pierwszej próbki, R_2 i $\overline{X}_2$ – rozstęp i średnia drugiej próbki, itd.

Ustalamy następnie skale dla obu torów karty: toru $\overline{X}$ i toru R. Pomocne będą przy tym podane poniżej wskazówki, choć w niektórych przypadkach i tak konieczne będą pewne modyfikacje przyjętych początkowo skal.

W przypadku wykresu $\overline{X}$ różnica na skali między wartościami największymi i najmniejszymi powinna być co najmniej 2 razy większa niż różnica między największą i najmniejszą wartością średniej próbek $\overline{X}_i$.

W przypadku wykresu R wartości powinny rozpoczynać się od zera, natomiast wartość górna skali powinna być 2 razy większa od największego rozstępu R.

Pewną wskazówką jest ustalenie obu skal w taki sposób, by jednostka na skali rozstępu była podwojeniem jednostki na skali średnich. Przykład: jeśli jednostka na skali $\overline{X}$ wynosi 0,001mm, to na skali R przyjmujemy jednostkę o wartości 0,002 mm. Przy tym układzie granice kontrolne na obu wykresach będą miały w przybliżeniu tę samą szerokość.

Gdy mamy już ustalone obie skale, rysujemy na części graficznej formularza karty kontrolnej na torze $\overline{X}$ linię poziomą na wysokości $\overline{\overline{X}}$ a na torze R linię poziomą na wysokości $\overline{R}$. Są to tzw. linie centralne.

Następnie na obu torach karty w punktach odpowiadających kolejnym próbkom „*i*" nanosimy wyliczone zgodnie z częścią kalkulacyjną formularza wartości parametrów statystycznych $\overline{X}_i$ i R_i. Otrzymane w ten sposób punkty łączymy kolejno linią łamaną.

Kartę kontrolną prowadzoną w trakcie badań wstępnych zaznaczamy w nagłówku adnotacją „Badania wstępne".

Obliczanie linii kontrolnych metodą stabilizacyjną

Następnym krokiem w wypełnianiu karty kontrolnej jest naniesienie na obu jej torach linii kontrolnych.

Zewnętrzne linie kontrolne (górną i dolną) dla toru X wyznacza się z poniższych wzorów:

$$\overline{X}_g = \overline{\overline{X}} + A_2\overline{R}; \qquad \overline{X}_d = \overline{\overline{X}} - A_2\overline{R}$$

Dla toru R zewnętrzne linie kontrolne (górną i dolną) określają wzory:

$$R_g = D_4\overline{R}; \qquad R_d = D_3\overline{R}$$

Dla próbek o liczebności mniejszej od 7 nie wyznacza się dolnej linii kontrolnej R_d.

W przypadkach, gdy nawet niewielkie zmiany w procesie produkcyjnym mają duże znaczenie dla poziomu jakości, na torach kart kontrolnych wyznacza się dodatkowo wewnętrzne linie kontrolne (linie ostrzegawcze) – górną i dolną. Przekroczenie tych linii jest sygnałem ostrzegawczym o zbliżającym się rozregulowaniu procesu lub, w przypadku kolejnego przekroczenia, sygnałem o rozregulowaniu. Linie ostrzegawcze umieszczone są w odległości wynoszącej około 0,65 odległości między liniami kontrolnymi. Położenie linii ostrzegawczych (wewnętrznych linii kontrolnych) można wyznaczyć posługując się poniższymi wzorami:

$$\overline{X}'_g = \overline{\overline{X}} + A'_2\overline{R}; \qquad \overline{X}'_d = \overline{\overline{X}} - A'_2\overline{R}; \qquad R'_g = D'_4\overline{R}$$

Wartości współczynników A_2, A'_2, D_3, D_4, D'_4 w zależności od liczebności próbki podano w poniższej tabeli:

n	A_2	A'_2	D_3	D_4	D'_4	d_n
2	1,880	1,229	0	3,27	2,46	1,128
3	1,023	0,663	0	2,57	1,96	1,693
4	0,729	0,476	0	2,28	1,76	2,059
5	0,577	0,377	0	2,11	1,66	2,326
6	0,483	0,316	0	2,00	1,59	2,534
7	0,419	0,274	0,076	1,92	1,54	2,704
8	0,373	0,243	0,136	1,86	1,51	2,842
9	0,337	0,220	0,184	1,82	1,48	2,970
10	0,308	0,201	0,223	1,78	1,45	3,078

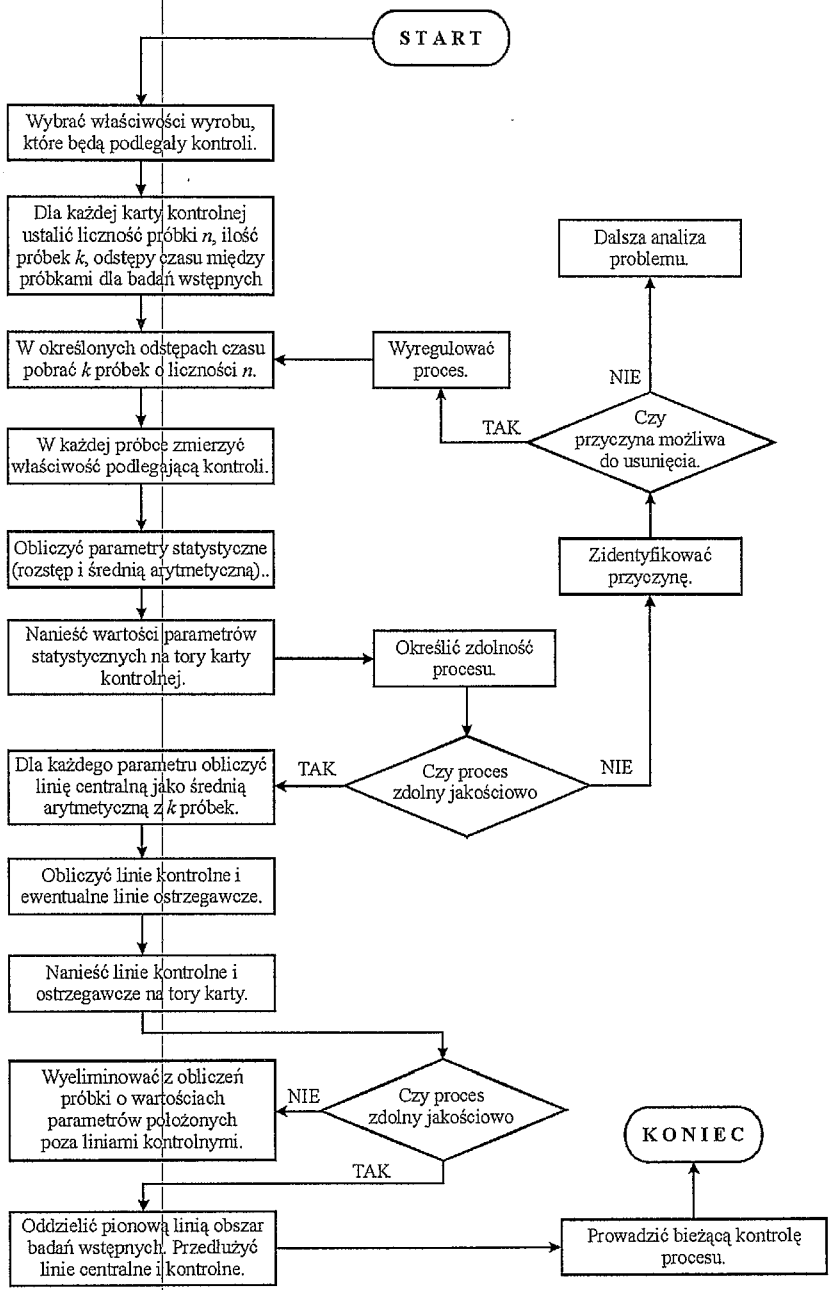
Gdy po wrysowaniu linii kontrolnych okazuje się, że niektóre punkty wykresu przekraczają je, to jest to dowodem niestabilności procesu. Konieczne jest wówczas przeprowadzenie analizy procesu w celu wykrycia przyczyn tych rozregulowań, a następnie ich wyeliminowania.

Z formularza wykreślamy punkty znajdujące się poza liniami kontrolnymi i powtarzamy obliczenia linii centralnych, kontrolnych i ostrzegawczych.

Bieżąca kontrola procesu

Po zakończeniu badań wstępnych oddzielamy na karcie $\overline{X} - R$ linią pionową wypełnioną dotychczas część formularza. Przedłużamy linie centralne i kontrolne i przystępujemy do bieżącej kontroli procesu.

Sposób prowadzenia karty $\overline{X} - R$ z obliczaniem linii kontrolnych metodą stabilizacyjną przedstawia algorytm pokazany na rysunku 1.



Rysunek 1
Algorytm postępowania dla metody stabilizacyjnej

Obliczanie linii kontrolnych metodą projektową

Metodę tę stosujemy wówczas, gdy projektuje się kartę kontrolną dla produkcji nowo uruchamianej, która nie daje w pierwszej fazie możliwości zebrania danych statystycznych.

Do obliczeń wykorzystuje się narzucone przez dokumentację konstrukcyjną granice tolerancji T_g i T_d oraz założony dopuszczalny poziom wadliwości w_2 (wg normy PN-ISO 3534-2, punkt 3.4.2 w_2 ma międzynarodowe oznaczenie AQL i nazwę: akceptowany poziom jakości).

Położenie linii kontrolnych i ostrzegawczych określają wzory:

$$\overline{X}_d = T_d + S \cdot T; \qquad \overline{X}_g = T_g - S \cdot T; \qquad R_g = H_2 \cdot T$$

$$\overline{X}'_d = T_d + 0{,}35 \cdot S \cdot T; \qquad \overline{X}'_g = T_g - 0{,}35 \cdot S \cdot T; \qquad R'_g = 0{,}65 \cdot H_2 \cdot T$$

Wartości współczynników S i H_2 w zależności od założonej wadliwości w_2, liczebności próbki i założonego prawdopodobieństwa przekroczenia linii kontrolnych p przedstawia poniższa tabela.

Wartości współczynników S i H_2

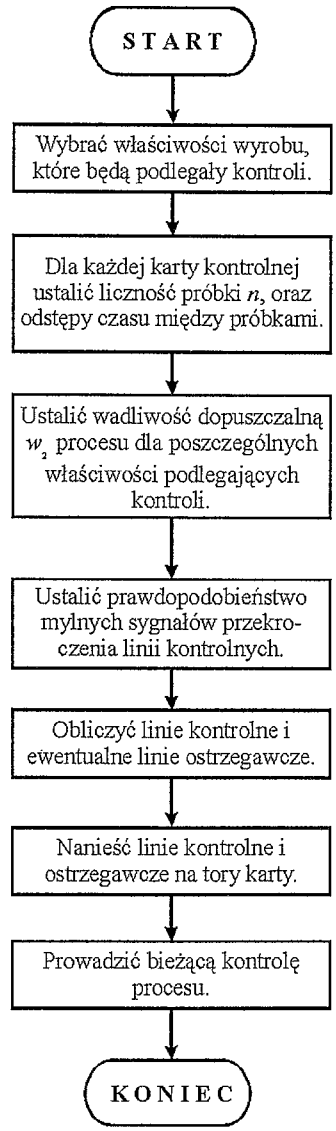
$p\overline{x}$	w %	2,0	1,0	0,5	0,27
	t_1	2,33	2,58	2,81	3,00
T	n	S			
1% 2,58	4	0,223	0,250	0,271	0,285
	5	0,252	0,276	0,295	0,308
	6	0,274	0,296	0,313	0,325
	10	0,325	0,342	0,355	0,364
0,5% 2,81	4	0,198	0,228	0,250	0,266
	5	0,230	0,256	0,276	0,291
	6	0,254	0,278	0,296	0,309
	10	0,309	0,328	0,342	0,352
0,27% 3,00	4	0,177	0,209	0,233	0,250
	5	0,211	0,240	0,261	0,276
	6	0,236	0,262	0,282	0,296
	10	0,296	0,316	0,331	0,342

pR_g	w %	2,0	1,0	0,5	0,27
	t_1	2,33	2,58	2,81	3,00
t	n	H_2			
1% 2,33	4	0,94	0,85	0,78	0,73
	5	0,99	0,89	0,82	0,77
	6	1,02	0,92	0,85	0,79
	10	1,11	1,00	0,92	0,86
0,5% 2,58	4	1,00	0,91	0,83	0,78
	5	1,05	0,95	0,87	0,81
	6	1,08	0,98	0,90	0,84
	10	1,16	1,05	0,96	0,90
0,25% 2,81	4	1,07	0,96	0,89	0,83
	5	1,10	1,00	0,92	0,86
	6	1,14	1,03	0,94	0,88
	10	1,22	1,10	1,01	0,95

W praktyce zazwyczaj kartę $\overline{X} - R$ prowadzi się z prawdopodobieństwem 0,27% na torze $\overline{X}$ i 0,5% na torze R, nie podaje danych do obliczania dolnej linii kontrolnej dla rozstępu, gdyż uważa się ją za zbędną. Zaletą metody projektowej jest narzucanie z góry wymagań jakościowych, jakie ma spełniać pro-

ces. Ponieważ początkowy okres prowadzenia karty tą metodą jest okresem stabilizacji procesu, wyprodukowane na tym etapie elementy powinny być kontrolowane w 100%.

Algorytm postępowania dla metody projektowej przedstawia rysunek 2.



Rysunek 2
Algorytm postępowania dla metody projektowej

Wnioski

1. Nowoczesna kontrola jakości nie może ograniczać się do badania zgodności wykonywanych produktów z dokumentacją konstrukcyjną wraz z warunkami technicznymi i eliminacji braków w celu niedopuszczenia do użytkowników produktów o niewystarczającej jakości i nieprzydatnych w eksploatacji.

2. Ma ona za zadanie stosowanie metod najbardziej ekonomicznych, tzn. koszty kontroli powinny być mniejsze niż koszt braków, które powstałyby, gdyby kontroli nie było (odstępstwa tylko w wyjątkowych przypadkach, np. gdy braki mogą zagrażać życiu ludzkiemu itp.).

3. Statystyczna kontrola jakości przy wykorzystaniu możliwości informatyki spełnia powyższe wymagania.

Literatura

WASILEWSKI L.: *Metody kontroli jakości w przedsiębiorstwach przemysłowych.*
HAMROL A., MANTURA W.: *Zarządzanie jakością, teoria i praktyka.*
KOWALSKI K.: *Statystyczne metody kontroli jakości.* Koszalin 2000.

Statistical quality control

Abstract

In the paper there is presented alghoritmisation of choosen methods using in quality control of ready products. There is very well known, that the goal for such control is improving the mass and multiserial productionj. Tbe base for this are quantitative mathematical statistical methods, including probability theory. Presented methodical proposals are using in special informatical systems which include also statistical quality control modul. This is a part of some statistical packages as for example STATISTICS by Statsoft Corporation.

Stanisław Stańko, Marcin Idzik
Katedra Ekonomiki Rolnictwa i Międzynarodowych
Stosunków Gospodarczych SGGW

Zastosowanie różnych metod ilościowych w prognozowaniu zjawisk i procesów gospodarczych na podstawie szeregów czasowych, w których występują tendencja, wahania sezonowe i przypadkowe

Wstęp

Obecnie w wielu dziedzinach nauk empirycznych szerokie zastosowanie znajduje modelowanie matematyczne. Na podstawie modelu matematycznego obserwuje się i sprawdza prawa rządzące zjawiskami, najczęściej po to, aby przewidzieć, jak te zjawiska będą przebiegać. Proces poznawania rzeczywistości za pomocą modelu matematycznego polega na budowie modelu rozpatrywanego zjawiska, statystycznej estymacji jego parametrów i wnioskowaniu o mechanizmie jego rozwoju w czasie.

Predykcja ekonometryczna z zastosowaniem modelu matematycznego ma charakter statystyczno-ekstrapolacyjny, czyli podstawą tego rodzaju wnioskowania w przyszłość jest przede wszystkim statystyczna obserwacja związków i prawidłowości występujących między zmiennymi ekonomicznymi. Jeżeli istnieją podstawy, aby przypuszczać, że występujące w przeszłości powiązania między zmiennymi będą również aktualne w przyszłości, to związki te rzutuje się w przyszłość i na ich podstawie buduje się prognozę.

Przedstawiona w pracy analiza metod prognozowania zjawisk, w których poziom zmiennej endogenicznej jest zdeterminowany oddziaływaniem składnika systematycznego, periodycznego oraz losowego, jest analizą opartą na danych w postaci szeregu czasowego i wykorzystującą formalne metody statystyczne, a wyniki prognoz zależą bezpośrednio od zastosowanych metod prognozowania.

W sytuacjach, gdy predykcja ma charakter średniookresowy, model ekonometryczny powinien uwzględniać nie tylko trend i wahania losowe zmiennej objaśnianej, ale także różne jej wahania typu periodycznego. W większości

przypadków w rolnictwie wahania periodyczne powtarzają się w okresach roku, ale ich amplituda zależy od czynników naturalnych, technicznych, instytucjonalnych i społecznych,

Przedstawione w pracy zagadnienia można podzielić na dwie części. Pierwsza z nich dotyczy budowy oraz weryfikacji modeli matematycznych dla analizowanego zagadnienia, druga przedstawia problemy związane z praktycznym ich wykorzystaniem do celów prognostycznych. W artykule dokonano konstrukcji oraz oceny prognoz dla danych kwartalnych z wykorzystaniem metod: dekompozycji szeregu czasowego, wyrównywania wykładniczego Wintersa, ekstrapolacji trendów jednoimiennych okresów, funkcji regresji oraz analizy harmonicznej. Przykład empiryczny dotyczył kształtowania się skupu mleka w Polsce w okresie od pierwszego kwartału 1993 r. do pierwszego kwartału 2000 r. Przyjęto założenia, że horyzont prognozy obejmował cztery kwartały. Moment konstrukcji prognozy wyznacza czwarty kwartał każdego roku, poczynając do 1994 do 1999 r. Takie postępowanie umożliwia stopniowe wydłużanie okresu retrospekcji oraz szacowanie uaktualnionych modeli prognostycznych, na podstawie większej ilości informacji. Zastosowanie analizy ex post w stosunku do uzyskanych wyników prognozowania umożliwiło ocenę przydatności wspomnianych metod w prognozowaniu zjawisk o charakterze sezonowym.

Metody prognozowania skupu mleka

W literaturze przedmiotu występuje wiele metod uwzględniających wpływ wahań periodycznych na zmienną prognozowaną (m.in. Analiza ... [1996], Cieślak [1997], Dittman [1996], Filipiak [1998], Farnum, Stanton [1989], Grabiński i inni [1990], Kalisiak [1978], Kildiszew [1976], Mlinowska-Wasyl [1970], Nowak [1998], Radzikowska [1999], Stańko [1999], Woźniak, Zeliaś [1975], Zeliaś [1997], Zieliński [1979]). W tym opracowaniu uwaga została zwrócona tylko na niektóre z nich.

Metoda Wintersa należy do adaptacyjnych technik prognozowania, zasługując na uwagę ze względu na fakt możliwości rozluźniania podstawowych założeń klasycznej teorii predykcji, o których piszą Pawłowski [1973], Stańko [1999], Zeliaś [1997]. Korzystanie z modelu Wintersa wymaga przyjęcia postawy pasywnej.

Ogólną postać modelu multiplikatywnego można opisać wzorem:

$$y_t = (\beta_0 + \beta_1 t)S_t + \varepsilon_t \qquad (1)$$

Prognozę w okresie (t) dla (p) okresów buduje się według wzoru:

$$\tilde{y}_{t+p}(t) = T^*_{t+p}(t) S^*_{t+p}(t) = \left[T^*_t(t) + p\beta_t(t)\right] S^*_{t+p}(t) \tag{2}$$

Stosowanie tej techniki prognozowania wymaga: danych empirycznych (y_t), oceny trendu $\left(T^*_t\right)$ w okresie t, oceny zmian trendu β^*_1 w okresie t i wahań sezonowych dla poszczególnych okresów S^*_t, obliczonych w okresie t. Szczegółowy opis tej metody można znaleźć w pracach: Cieślak [1997], Dittman [1996], Farnum i Stanton [1989], Radzikowska [1999], Stańko [1999].

Do predykcji zmiennej prognozowanej na podstawie funkcji trendu z wahaniami periodycznymi można wykorzystać także metodę wskaźników, nazywaną dekompozycją sezonową. Szczegółową charakterystykę tej metody przedstawiają: Cieślak [1996], Kildiszew [1976], Radzikowska [1999], Stańko [1999], Zając [1994], natomiast modyfikację tej metody opartą na iteracyjnym podejściu prezentuje Dittman [1996]. Badania wielkości wahań sezonowych dokonujemy metodą klasycznej dekompozycji sezonowej Census I. Prognozę buduje się przez ekstrapolację funkcji trendu i uwzględnienie efektu wahań sezonowych:

$$\tilde{y}_{t+p}(t) = \tilde{y}_{t+p} \times S_p \tag{3}$$

gdzie: $\tilde{y}_{t+p}(t)$ – prognoza zmiennej prognozowanej w okresie t,

$\tilde{y}_{t+p}$ – prognoza trendu w okresie t,

S_p – czysty wskaźnik sezonowości.

Dla szeregu czasowego, w którym występują tendencja, wahania sezonowe i przypadkowe, do konstrukcji prognozy można także zastosować metodę ekstrapolacji trendów jednoimiennych okresów, według której modele tendencji rozwojowej budowane są oddzielnie dla poszczególnych okresów. Zbiór zaobserwowanych wartości zmiennej zależnej ($Y_{h=mk}$, h = 1, . . . , m, k = 0, 1, . . . , n – 1) dzielimy na *m* szeregów czasowych odnoszących się do tego samego okresu (sezonu) (Radzikowska [1999], Zeliaś [1997]). Poszczególne szeregi czasowe opisujemy funkcją trendu, która w przypadku modelu liniowego ma postać:

$$Y_{kh} = \beta_{oh} + \beta_{1h} t_{kh} + \varepsilon_{kh} \tag{4}$$

gdzie: Y_{kh} – wartości szeregu czasowego w okresie (sezonie) h,

t_{kh} – zmienna czasowa t taka, że $t_{kh} = h + m(k - 1)$,

β_{oh}, β_{1h} – parametry strukturalne h-tego modelu,

ε_{kh} – składnik losowy.

Parametry modelu szacujemy za pomocą klasycznej metody najmniejszych kwadratów. W przypadku modelu liniowego otrzymujemy:

$$Y_{kh} = \beta^*_{oh} + \beta^*_{1h} t_{kh} + \varepsilon_{kh} \tag{5}$$

gdzie: $\beta^*_{oh}, \beta^*_{1h}$ są ocenami parametrów β_{oh}, β_{1h} wyznaczonymi na podstawie n obserwacji [Grabiński i inni 1981]. Rozwiniętą koncepcję wyznaczania trendów jednoimiennych okresów przedstawiła Malinowska-Wasyl [1970].
Równanie prognozy ma postać:

$$\tilde{y}_{Th} = \beta^*_{oh} + \beta^*_{1h} T, \; T = n + 1, h = 1, ..., m \tag{6}$$

Inny sposób uwzględniania wahań periodycznych polega na przedstawieniu ich w postaci analizy harmonicznej. Charakterystykę tej metody można znaleźć w pracach: Radzikowska [1999], Cieślak [1997], Dittman [1996], Kildiszew [1976]. Analiza harmoniczna polega na budowie tzw. harmonik, tj. funkcji sinusoidalnych lub cosinusoidalnych o danym okresie. Pierwsza harmonika ma okres równy długości okresu badanego, druga – połowę tego okresu, trzecia – jednej trzeciej tego okresu itd. W przypadku n obserwacji liczba wszystkich harmonik jest równa $\frac{n}{2}$. Zapis tego modelu jest następujący:

$$y = \sum_{i=0}^{p} \alpha_i t^i + \sum_{j=1}^{\frac{m}{2}} \beta_{1j} \cos \omega_j t + \sum_{j=1}^{\frac{m}{2}} \beta_{2j} \sin \omega_j t + \varepsilon_t \tag{7}$$

gdzie: $\omega_j = \frac{2j\Pi}{m}$, $j = 1, \ldots, \frac{m}{2}$, $t = 1, \ldots, n$

$y = (y_1, \ldots, y_n)$, $\alpha = (\alpha_0, \alpha_1, \ldots \alpha_p)$, $\beta_1 = (\beta_{11}, \ldots, \beta_{1\frac{m}{2}})$,

$\beta_2 = (\beta_{22}, \ldots, \beta_{2\frac{m}{2}-1})$, $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$

Postać prognozy wynika z modelu:

$$\tilde{y}_t = f(t) + \sum_{j=1}^{\frac{m}{2}} \left[\beta_{1j} \cos \frac{2j\pi}{m} + \beta_{2j} \sin \frac{2j\pi}{m} \right] \qquad (8)$$

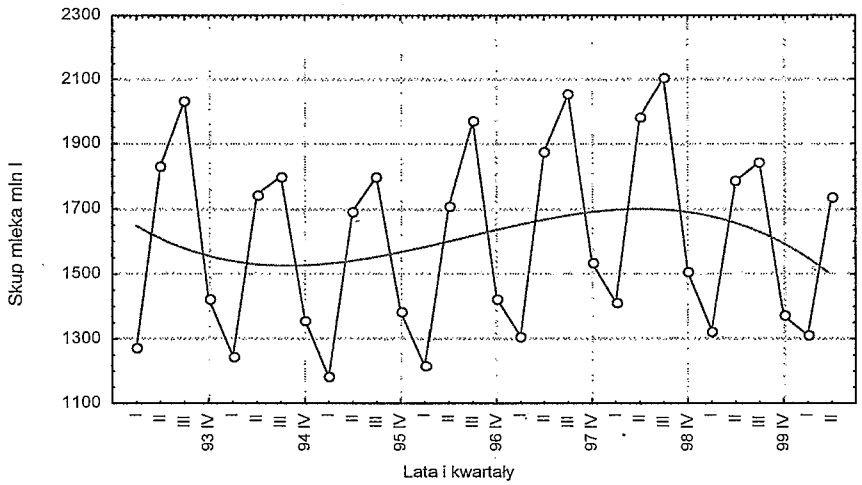
Do budowy prognozy na podstawie szeregu czasowego z wahaniami sezonowymi można także wykorzystać modele funkcji regresji (Farnum, Stanton [1989], Stańko [1999], Jóźwiak [2000]). W przypadku modelu multiplikatywnego postać funkcji regresji jest następująca:

$$Y_t = \beta_0 \times \beta_1^t \times S_1 x_1 \times S_2 x_2 \times \ldots \times S_L x_L \times \varepsilon \qquad (9)$$

gdzie: $x_1, x_2, \ldots x_L$ – zmienne dla okresów,
β_0, β_1 – parametry strukturalne funkcji,
ε_t – składnik wahań przypadkowych,
S – sezonowość.

Modele kształtowania się wielkości skupu mleka w Polsce

Skup mleka w Polsce charakteryzuje się sezonowością (wykres 1). Rosnącą ilość skupowanego mleka odnotowuje się w drugich, a najwyższą w trzecich kwartałach (23% wyższy niż średni skup w roku). Najniższy skup ma miejsce w pierwszych kwartałach i jest niższy o 22% w stosunku do średniego poziomu w roku. Opisując szereg czasowy skupu mleka w okresie od pierwszego kwartału 1993 do drugiego kwartału 2000 r. wielomianem stopnia trzeciego: $y = 1750{,}3 - 76{,}23 \times x + 7{,}055 \times x^2 - 0{,}166 \times x^3 + \varepsilon$, możliwe jest określenie wyraźnej tendencji spadkowej na początku oraz na końcu badanego okresu. Tendencja rosnąca występuje natomiast w środku badanego okresu. Powyższe spostrzeżenie uwzględnione w analizie dodatkowo wskaże na metody, których algorytm pozwala uwzględnić tego typu zmiany. Budowane prognozy będą prognozami średniookresowymi, ponieważ dotyczą roku n + 1, z uwzględnieniem wszystkich okresów.



Wykres 1
Kwartalny skup mleka w Polsce w latach 1993–2000, wg GUS

Prognozy uzyskane metodą Wintersa

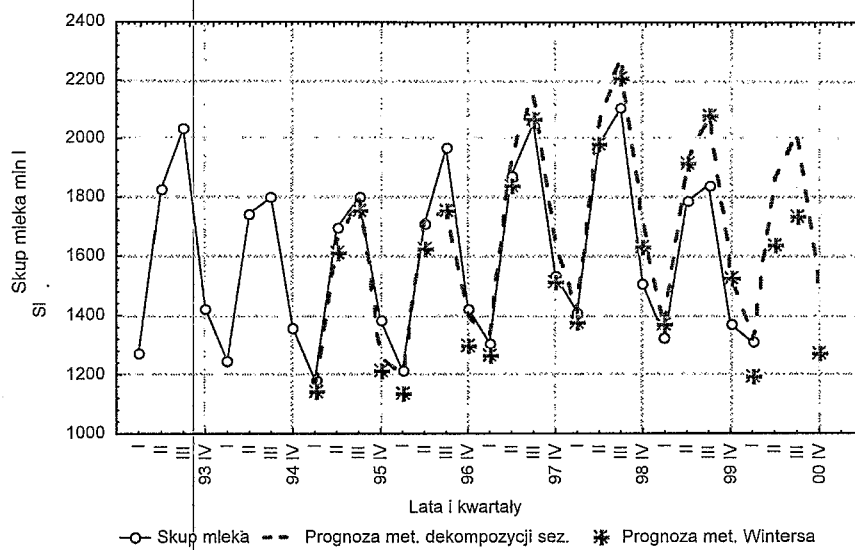
Do konstrukcji prognoz zastosowano model multiplikatywny Wintersa. Przy wyborze parametrów wygładzania zastosowano kryterium minimalizacji średniego kwadratowego błędu prognoz ex post ([Dittman 1996], [Zeliaś 1996], [Stańko 1999]). Stałe wygładzania oraz parametry modelu szacowano sekwencyjnie dla każdego roku, zwiększając jednocześnie stopniowo okres retrospekcji. Do wyznaczenia prognoz na rok 1995 oszacowano model na podstawie danych z lat 1993–1994, prognozy na rok 1996 uzyskano z modelu oszacowanego na podstawie danych z lat 1993–1995 itd.

Należy zauważyć, iż w dziesięciu przypadkach (tabela 1, wykres 4) otrzymano prognozy niedopuszczalne, przy przyjęciu rzędu dokładności prognoz w wysokości 5%. W pięciu przypadkach otrzymano prognozy zaniżone o więcej niż 5%, również tyle samo prognoz było niedopuszczalnie zawyżonych. Podczas rozpoczęcia tendencji wzrostowej w roku 1995 otrzymano prognozy zaniżone o 5,80%, tj. średnio o 99,92 mln l w skali roku. Podobna sytuacja miała miejsce w 1996 r. (tabela 2). Stosunkowo trafne prognozy uzyskano w 1997 r., gdy tendencja wzrostowa była kontynuowana. Wówczas bezwzględny błąd procentowy wyniósł 1,74%, tj. 28,70 mln l w skali roku. Gdy tendencja wzrostowa uległa zahamowaniu w 1998 roku, to prognozy skonstruowane na

ten rok były zawyżone średnio o 2,89%, przy błędzie bezwzględnym 4,06%, tj. 85,39 mln l w skali roku. W wyniku rozpoczęcia spadkowej tendencji skupu mleka w 1999 r. otrzymano prognozy zawyżone o 8,86%, tj. 159,18 mln l mleka rocznie. Ogólnie w okresie od pierwszego kwartału 1995 r. do pierwszego kwartału 2000 r. prognozy były zaniżone przeciętnie o 1,01%, a błąd bezwzględny procentowy prognoz wyniósł 6,13%, co stanowi średnio 111,43 mln l mleka w skali rozpatrywanego okresu.

Prognozy uzyskane metodą dekompozycji elementów szeregu czasowego

Do określenia wielkości wahań sezonowych zastosowano metodę klasycznej dekompozycji sezonowej Census I, model multiplikatywny. Postacie analityczne funkcji trendu wykorzystywanych do ekstrapolacji tendencji szacowane były sekwencyjnie w momencie konstrukcji prognozy dla każdego roku, podobnie jak w przypadku modelu Wintersa.



Wykres 2
Skup mleka i jego prognoza z zastosowaniem metody dekompozycji szeregu czasowego oraz metody Wintersa, model multiplikatywny

Przewidywane wielkości skupu mleka w poszczególnych latach należy uznać za zadowalające (wykres 2, tabela 1). Porównując otrzymane wielkości względnych, jak i bezwzględnych błędów prognozy z wartością krytyczną $\eta = 5\%$, możemy uznać wyznaczone prognozy, z wyjątkiem prognoz sporządzonych na rok 1998 ($\eta = 6{,}61\%$) za dopuszczalne (wykres 4, tabela 2). Stosunkowo trafne prognozy uzyskano w 1995 r., ich wyniki były zaniżone średnio o 2,97%, przy bezwzględnym błędzie procentowym 3,07%, tj. 66,07 mln l w skali roku. Podobna sytuacja miała miejsce w 1996 r., gdy prognozy zaniżone były o 3,08%, tj. średnio o 111,54 mln l w skali roku. W 1997 r., gdy tendencja wzrostowa była kontynuowana, prognozy były zawyżone to 4,52%, tj. średnio o 85,06 mln l w skali roku. Gdy tendencja wzrostowa uległa zahamowaniu w 1998 r., prognozy skonstruowane na ten rok były zawyżone średnio o 6,61%, przy błędzie bezwzględnym 6,69%, tj. 147,03 mln l w skali roku. W wyniku rozpoczęcia spadkowej tendencji skupu mleka w 1999 r. otrzymano prognozy zawyżone o 8,80%, tj średnio o 162,13 mln l w stosunku do rzeczywistego poziomu skupu mleka w Polsce w 1999 r. Ogólnie w okresie od pierwszego kwartału 1995 r. do pierwszego kwartału 2000 r. prognozy były zawyżone o 2,68%, a błąd bezwzględny prognoz wyniósł 4,4%, co stanowi średnio 117,20 mln l mleka w skali pięciu lat. Należy zauważyć, iż w ośmiu przypadkach otrzymano prognozy niedopuszczalne w świetle przyjętego 5-procentowego błędu.

Prognozy uzyskane metodą ekstrapolacji trendów jednoimiennych okresów

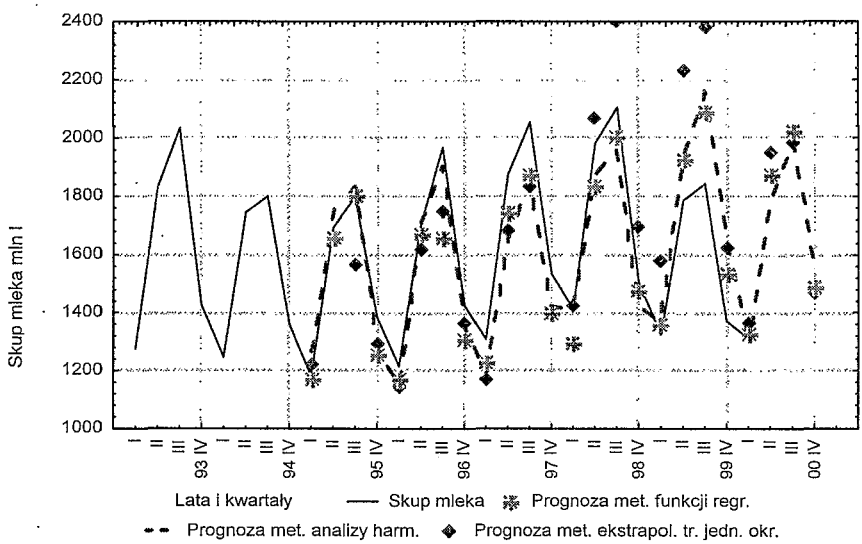
Korzystając z metody ekstrapolacji trendów jednoimiennych okresów, skonstruowano prognozy wielkości skupu mleka na poszczególne kwartały kolejnych lat ($t = k + 1$) (wykres 3). W tabeli 1 przedstawiono wartości błędów średnich oraz względnych procentowych dla prognoz w kolejnych kwartałach w latach 1995–2000. Rezultaty prognozowania z zastosowaniem metody ekstrapolacji trendów jednoimiennych okresów należy uznać za niezadowalające. W piętnastu przypadkach na 21 postawionych prognoz (wykres 6) otrzymano prognozy niedopuszczalne, z czego w dziewięciu przypadkach prognozy wskazywały wysoce zawyżone wartości w stosunku do rzeczywistego skupu. Ogólnie w okresie od pierwszego kwartału 1995 r. do pierwszego kwartału 2000 r. prognozy były zaniżone o 1,99%, a bezwzględny błąd procentowy prognoz wyniósł 10,27%, co stanowi średnio 218,47 mln l mleka w skali rozpatrywanego okresu (tabela 2).

Prognozy uzyskane metodą analizy harmonicznej

Przed oszacowaniem parametrów harmonik szereg czasowy został poddany procesowi detrendyzacji. Największy udział w wyjaśnianiu sezonowej zmienności skupu mleka w Polsce ma harmonika trzecia o cyklu równym 4 kwartały, wyjaśniając 88% zmienności zmiennej prognozowanej. Harmonika druga, prezentująca wahania o cyklu 1,5-rocznym, wyjaśniła 3% zmienności zmiennej. Ponieważ harmoniki druga i trzecia wyjaśniają w sumie 92% zmienności skupu mleka, do prognozowania wykorzystano model tylko z tymi harmonikami. Oszacowany model ma postać:

$$Y = f(t) + 0{,}421\sin(t\pi/3) - 76{,}40\cos(t\pi/3) - 328{,}8\sin(t\pi/2) - 175\cos(t\pi/2)$$

Wyniki prognozowania oraz błędy prognoz wyznaczonych metodą analizy harmonicznej w poszczególnych kwartałach przedstawiają tabela 1 oraz wykresy 3 i 5. W dwunastu przypadkach otrzymano prognozy niedopuszczalne w świetle maksymalnego 5-procentowego błędu. Podczas tendencji wzrostowej w 1995 roku otrzymano prognozy zawyżone o 1,50%. Błąd bezwzględny tych prognoz wyniósł 5,05%, co stanowi średnio odchylenie 71,54 mln l w skali roku. Trafne prognozy uzyskano w 1996 r., gdy były one zaniżone o 3,39%, przy błędzie bezwzględnym 3,6%, co stanowi 136,02 mln l w skali roku. W 1997 r. prognozy wskazywały, iż skup mleka będzie niższy niż to miało miejsce w rzeczywistości o 9,48%, tj. 163,21 mln l, podobnie jak w 1998 r. W wyniku postępującej spadkowej tendencji skupu mleka w 1999 r. otrzymano prognozy zawyżone średnio o 12,13%, tj. 195,93 mln l mleka rocznie. Ogólnie w rozpatrywanym okresie prognozy były zaniżone o 0,70%, a bezwzględny błąd procentowy prognoz wyniósł 5,83%, co stanowi średnio 108,25 mln l.



Wykres 3
Skup i prognoza skupu mleka z zastosowaniem metody analizy harmonicznej, metody ekstrapolacji trendów jednoimiennych okresów oraz funkcji regresji

Prognozy uzyskane metodą funkcji regresji

Korzystając z metody funkcji regresji skonstruowano prognozy skupu mleka na poszczególne kwartały kolejnych lat (wykres 3). Ogólnie w okresie od pierwszego kwartału 1995 r. do pierwszego kwartału 2000 r. prognozy wskazywały, iż skup mleka będzie niższy od rzeczywistego o 2,83%, przy bezwzględnym błędzie procentowym 5,62% (105,0 mln l) średnio w skali badanego okresu. W czternastu przypadkach na 21 postawionych prognoz otrzymano prognozy niedopuszczalne (tabela 1). Wyniki jedenastu prognoz były zaniżone o więcej niż 5%. Podczas tendencji wzrostowej w roku 1995 otrzymano prognozy zaniżone o 3,06%, podobnie jak w 1996 r. o 7,57% oraz w 1997 r. o 7,68%. Gdy tendencja wzrostowa uległa zwolnieniu w 1998 roku, prognozy skonstruowane na ten rok były zaniżone o 5,63%, tj. średnio 98,83 mln l w skali roku. W wyniku postępującej spadkowej tendencji skupu mleka w 1999 r. otrzymano prognozy zawyżone średnio o 8,87%, tj. 144,93 mln l mleka rocznie.

Tabela 1
Błędy prognoz przy zastosowaniu różnych metod prognozowania

Data obserwacji, rok i kwartał	Rzeczywisty skup mleka mln l	Błędy prognoz wg zastosowanych metod prognozowania									
		metoda Wintersa		dekompozycja sezonowa		ekstrapolacja trendów jednoimiennych okresów		analiza harmoniczna		funkcja regresji	
		RMSE*	MPE**	RMSE*	MPE**	RMSE*	MPE**	RMSE*	MPE**	RMSE*	MPE**
I 1995	1182	39,99	–3,38	9,82	–0,83	37	3,13	84,04	7,11	10,12	–0,86
II	1694	84,11	–4,97	35,34	–2,09	35	–2,07	67,97	4,01	36,83	–2,17
III	1798	44,77	–2,49	3,7	0,21	233	–12,96	35,71	1,99	0,4	0,02
IV	1385	171,04	–12,35	126,89	–9,16	92	–6,64	98,43	–7,11	127,88	–9,23
I 1996	1216	81,19	–6,68	12,25	–1,01	74,33	–6,11	72,17	–5,94	43,58	–3,58
II	1707	81,91	–4,8	28,29	1,66	86,33	–5,06	7,23	0,42	38,75	–2,27
III	1968	210,87	–10,71	219,46	–11,15	220,03	–11,18	63,88	–3,25	312	–15,85
IV	1425	127,43	–8,94	25,65	–1,8	63	–4,42	68,45	–4,8	122,08	–8,57
I 1997	1308	37,8	–2,89	24,96	1,91	136	–10,4	117,66	–9	80,43	–6,15
II	1875	32,01	–1,71	70,36	3,75	192,05	–10,24	222,43	–11,86	129,95	–6,93
III	2053	17,11	0,83	98,51	4,8	216,56	–10,55	202,51	–9,86	177,32	–8,64
IV	1535	23,43	–1,53	116,9	7,62	145	–9,45	110,25	–7,18	138,23	–9
I 1998	1412	33,23	–2,35	2,53	–0,18	13,6	0,96	9,9	–0,7	117,7	–8,34
II	1982	3,54	0,18	63,42	3,2	88,8	4,48	110,7	–5,59	146,07	–7,37
III	2104	109,22	5,19	180,3	8,57	293,4	13,94	140,07	–6,66	102,68	–4,88
IV	1506	128,41	8,53	223,47	14,84	188,6	12,52	75,23	–5	28,88	–1,92
I 1999	1326	46,8	3,53	31,17	2,35	252,2	19,02	37,67	2,84	32,48	2,45
II	1785	132,19	7,41	140,06	7,85	442,6	24,8	164,63	9,22	141,54	7,93
III	1843	237,42	12,88	240,26	13,04	539,7	29,28	319	17,31	241,41	13,1
IV	1371	159,13	11,61	163,79	11,95	252,3	18,4	262,42	19,14	164,28	11,98
I 2000	1311	112,41	–8,57	9,76	0,74	56	4,27	2,98	0,23	12,48	0,95

Źródło: Obliczenia własne. * odchylenie standardowe, ** średni błąd procentowy.

Podsumowanie

Przeprowadzone badania empiryczne i ich wyniki dla zjawisk z różną dynamiką zmian i wahaniami sezonowymi przyniosły znaczną liczbę interesujących spostrzeżeń.

W tabeli 2 zostały zamieszczone wyniki średnich rocznych względnych oraz bezwzględnych błędów prognoz skupu mleka w Polsce. Wartości błędów reprezentują uśrednione kwartalne błędy prognoz w kolejnych latach. Odchylenia prognoz generowanych przez poszczególne modele różnią się niekiedy znacznie, ale wszystkie prognozy wskazują na wzrost skupu mleka w latach 1996–1997. Odmienna sytuacja ma miejsce w latach 1998–1999, gdy w wyniku zmiany kierunku ogólnej tendencji rozwojowej wyniki uzyskanych prognoz znacznie różnią się od siebie ze względu na zastosowaną metodę prognozowania.

Tabela 2
Średnie roczne względne oraz bezwzględne błędy procentowe prognoz wg zastosowanych metod

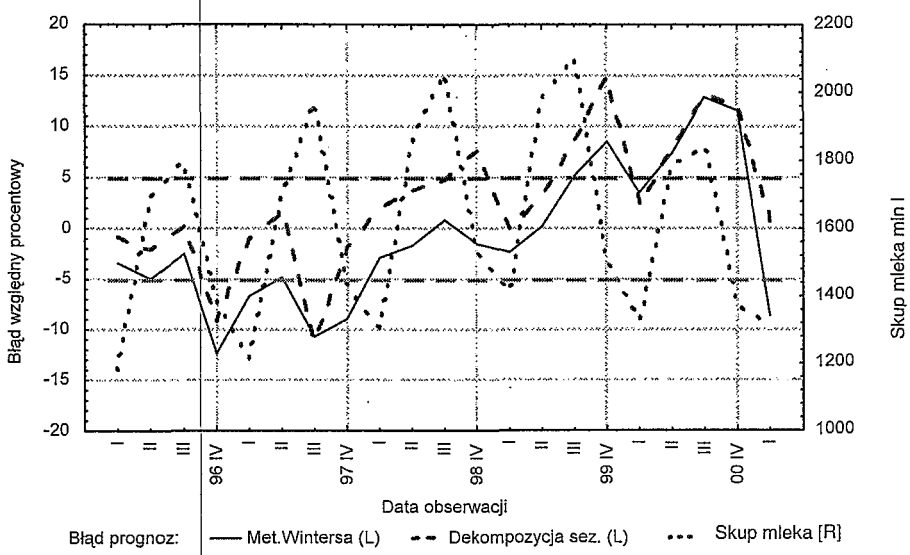
Lata	Metoda prognozowania									
	wygładzania wykładniczego Wintersa		dekompozycja sezonowa		ekstrapolacja trendów jednoimiennych okresów		analiza harmoniczna		funkcja regresji	
	Błąd procentowy									
	MPE	MAPE	MPE	MAPE	MPE	MAPE	MPE	MAPE	MPE	MAPE
1995	–5,80	5,80	–2,97	3,07	–4,63	6,20	1,50	5,05	–3,06	3,06
1996	–7,78	7,78	–3,08	3,09	–6,69	6,69	–3,39	3,60	–7,57	7,57
1997	–1,32	1,74	4,52	4,52	–10,16	10,16	–9,48	9,48	–7,68	7,68
1998	2,89	4,06	6,61	6,69	7,98	7,98	–4,48	4,48	–5,63	5,63
1999	8,56	8,56	8,80	8,80	22,88	22,88	12,13	12,13	8,87	8,87
2000	–8,57	8,57	0,74	0,74	4,27	4,27	0,23	0,23	0,95	0,95
Średnio	–1,01	6,13	2,68	4,40	1,99	10,27	–0,70	5,83	–2,83	5,62

Źródło: Obliczenia własne. MPE – średni błąd procentowy, MAPE – średni bezwzględny błąd procentowy prognoz.

Rozpatrując skuteczność prognozowania z zastosowaniem omawianych metod, należy stwierdzić, iż praktyczne zastosowanie prognostyczne w omawianym przypadku może znaleźć metoda dekompozycji szeregu czasowego. Jedynie w przypadku tej metody średni bezwzględny błąd procentowy progno-

zy w okresie od pierwszego kwartału 1995 r. do pierwszego kwartału 2000 r. nie przekracza wartości krytycznej (η = 5%), który najczęściej przyjmuje się w prognozach w rolnictwie.

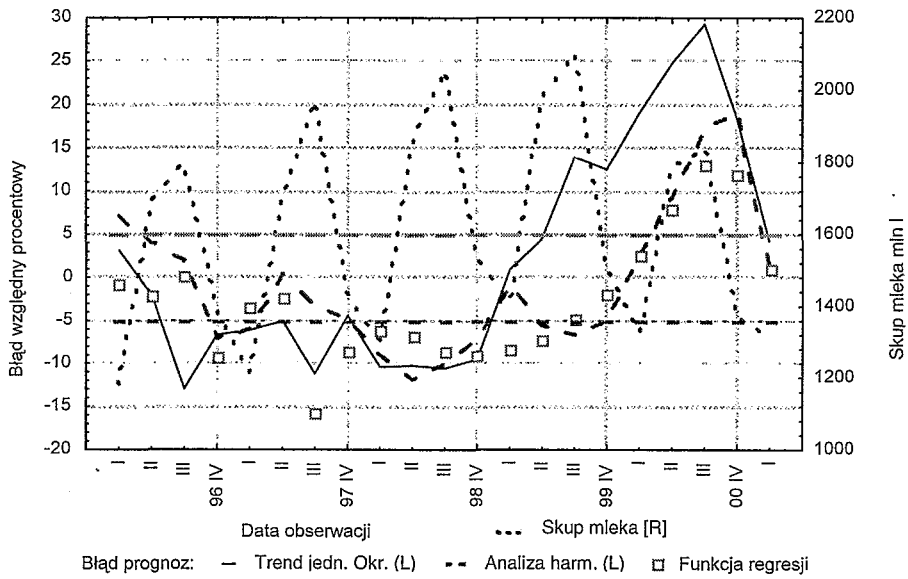
Charakteryzując prezentowane techniki prognostyczne w ocenie Steinhausa [1956] pod kątem wartości informacyjnej, poznawczej oraz ostrzegawczej, należy wskazać na metodę dekompozycji szeregu czasowego oraz funkcji regresji, jako te, które dają stosunkowo najdokładniejsze prognozy. Trafna prognoza w tym przypadku zależy głównie od poprawnie oszacowanego trendu. Na uwagę zasługuje natomiast prostota konstrukcji modelu, łatwa interpretacja oraz wysoka wartość poznawcza wyodrębnionych pośrednio elementów składowych modelu, których uzyskanie jest niejednokrotnie celem głównym stosowania różnych technik analizy szeregów czasowych. Warto nadmienić, że modele funkcji regresji były szacowane na podstawie danych nie oczyszczonych z wahań sezonowych, gdyż często po wyeliminowaniu wahań sezonowych traci się możliwość uzyskania dobrego modelu, generującego wielkości przez model odzwierciedlane w sposób prawidłowy. Podczas ustabilizowanej tendencji błędy prognoz w niewielkim stopniu przekraczały wartość krytyczną błędu (wykres 4).



Wykres 4
Analiza ex post błędów bezwzględnych prognoz uzyskanych metodą Wintersa oraz dekompozycji szeregu czasowego

Prognozy uzyskane na podstawie metody dekompozycji szeregu czasowego można uznać za wiarygodne, zważywszy na niewielkie wartości błędów prognoz. Podstawą takiego wyniku prognozowania jest możliwość odrębnego modelowania poszczególnych elementów szeregu, co w powiązaniu z doświadczeniem prognosty daje zadowalające rezultaty. Wyniki prognoz uzyskane innymi z omawianych metod charakteryzują się większymi błędami, jednak nie powinno to powodować rezygnacji z praktycznego zastosowania tych metod w procesie predykcji innych zjawisk gospodarczych. Dokładność prognozy zależy od typu modelu i jego właściwości.

Zastosowanie modeli do prognozowania w rolnictwie nie może odbywać się w sposób automatyczny ze względu na duży wpływ czynników losowych i instytucjonalnych na wyniki uzyskiwanych prognoz. Poprawna konstrukcja prognozy wymaga prognozowania ogólnej tendencji rozwojowej i odrębnego prognozowania odchyleń od tej tendencji, wywołanych działaniem między innymi czynników przyrodniczych.



Wykres 5
Analiza ex post błędów bezwzględnych prognoz

Najważniejszy wniosek praktyczny, jaki wynika z przedstawionej w pracy analizy porównania wartości prognostycznej niektórych popularnych i często stosowanych metod prognozowania zjawisk o charakterze sezonowym, polega na tym, że z punktu widzenia decydenta i prognosty ważne jest, aby zastosowa-

na technika dostarczała jak największej liczby informacji o prognozowanym zjawisku, pozwalała poznać strukturę szeregu czasowego i wskazać momenty, w których kształtujący się proces ulega zmianie.

Pomimo powszechnie dostępnych zaawansowanych obliczeniowych technik komputerowych, głównym zadaniem w procesach prognostycznych jest uzyskanie trafnych prognoz możliwie niskimi kosztami, co związane jest z dostępnością danych oraz czasem potrzebnym do uzyskania wyniku. Dlatego w praktycznym prognozowaniu należy głównie zwrócić uwagę na wiedzę i doświadczenie prognosty, a nie pokładać stuprocentowych nadziei uzyskania trafnej prognozy tylko przez mechaniczne zastosowanie bardziej zaawansowanej metody prognostycznej.

Literatura

Analiza stosowalności zagranicznych metod prognozowania plonów w warunkach Polski. IUNiG, Puławy 1996.

DITTMAN P.: *Metody prognozowania sprzedaży w przedsiębiorstwie.* AE, Wrocław 1996.

FARNUM N.R., STANTON W.: *Quantitative Forecasting Methods.* PSW-Kent Publishing Company, 1989.

FILIPIAK K.: *Modele stochastyczne.* IUNG, Puławy 1998.

GRABIŃSKI T., MALINA A., ZELIAŚ A.: *Metody analizy danych na podstawie szeregów przekrojowo-czasowych.* AE, Kraków 1990.

JÓŹWIAK J., PODGÓRSKI J.: *Statystyka od podstaw.* PWE, Warszawa 2000.

KALISIAK: *Ekonometryczne prognozy handlu zagranicznego.* PWN, Warszawa 1978.

KILDISZEW G., FRENKEL A.: *Analiza szeregów czasowych i prognozowanie.* PWE, Warszawa 1976.

MALINOWSKA-WASYL M.: *Budowa prognozy statystycznej dla procesu z wahaniami sezonowymi. Wybrane problemy prognoz statystycznych.* Warszawa 1970.

NOWAK E.: *Prognozowanie gospodarcze.* Placet, Warszawa 1998.

PAWŁOWSKI Z.: *Prognozy ekonometryczne.* PWN, Warszawa 1973.

Prognozowanie gospodarcze. Metody i zastosowania. Red. M. Cieślak. Wydawnictwo Naukowe PWN, Warszawa 1997.

RADZIKOWSKA B.: *Metody prognozowania.* AE, Wrocław 1999.

STAŃKO S.: Prognozowanie w rolnictwie. Wydawnictwo SGGW, Warszawa 1999.

STEINHAUS H.: O prognozie. *Zastosowania Matematyki,* z. 1, 1956.
WOŹNIAK M., ZELIAŚ A.: O metodach badania dynamicznej struktury szeregów czasowych. *Przegląd Statystyczny* 1975, nr 2.
ZAJĄC K.: *Zarys metod statystycznych.* PWE, Warszawa 1994.
ZELIAŚ A.: *Teoria prognozy.* PWE, Warszawa 1997.
ZIELIŃSKI Z.: *Metody analizy dynamiki rytmiczności zjawisk gospodarczych.* PWE, Warszawa 1979.

The application of some quantitative forecasting methods for economic time series with trend, seasonal and irregular components

Abstract

The article presents some useful methods for forecasting economical time series with trend and seasonal component. The example concerns milk time series for years 1991–1999. The predictions was made by the following methods: Winter's exponential smoothing, seasonal decomposition, harmonic analysis, regresion analysis, extrapolating a trend the same stage. The authors presented the checking of the methods for the years 1993–1999, which gave satisfactory results – a right forecast highly depends on forecasting methods.

Monika Krawiec
Instytut Ekonomii i Zarządzania Politechniki Szczecińskiej

Konstrukcja i wycena opcji na towarowy kontrakt futures w warunkach polskiego rynku*

Wstęp

Lata osiemdziesiąte i dziewięćdziesiąte XX wieku to okres znaczącego wzrostu wykorzystania rozmaitych instrumentów finansowych, tworzonych w celu wspomagania zarządzania ryzykiem. Do grupy takich instrumentów zalicza się kontrakty forward, futures, swap oraz opcje. Nie są to nowe narzędzia, z wyjątkiem kontraktów swap, gdyż zarówno opcje, jak i forwards oraz futures były wykorzystywane w różnych formach już w średniowieczu. Jednakże to w latach siedemdziesiątych XX wieku nastąpił wzrost znaczenia rynku instrumentów pochodnych. Przyczyniło się do tego m.in. zwiększenie płynności rynków, będące efektem przemian światowej gospodarki, jakie nastąpiły po II wojnie światowej.

Wiele metod i modeli, sformułowanych pierwotnie na potrzeby inżynierii finansowej, można zastosować także na rynkach towarowych. Na światowych giełdach towarowych dzieje się tak od wielu lat. W Polsce natomiast towarowe rynki terminowe znajdują się na razie w początkowej fazie rozwoju. Na dwóch polskich giełdach towarowych, prowadzących handel derywatami, to jest na Giełdzie Poznańskiej (GP) i Warszawskiej Giełdzie Towarowej (WGT), większość obrotów przypada na gotówkowe transakcje natychmiastowe.

Ponadto, wadą polskiego rynku terminowego jest niewielki asortyment dostępnych instrumentów, podczas gdy na przykład na amerykańskich giełdach notowane są tysiące kontraktów i wciąż wprowadza się nowe. Rozwojowi rynku nie sprzyja także niski poziom wiedzy potencjalnych inwestorów na temat instrumentów pochodnych i sposobów ich wykorzystania. Częściowo wynika to z pewnością ze słabej aktywności edukacyjno-informacyjnej zarówno uczelni, związanych programowo z produkcją, obrotem, towaroznawstwem i ekonomiką produkcji towarów rolnych, jak i samych giełd towarowych [5]. Po-

*Źródło finansowania: Komitet Badań Naukowych – RKMD/2000.

ważną niedogodnością jest brak spójnego systemu informacyjnego, a dane dostępne na internetowych stronach Giełdy Poznańskiej i Warszawskiej Giełdy Towarowej są niepełne.

Należy się jednak spodziewać dalszego rozwoju rynku derywatów w Polsce, prowadzącego do wykorzystywania coraz bardziej złożonych instrumentów finansowych. Stąd celem niniejszej pracy jest przedstawienie propozycji wprowadzenia instrumentu, niedostępnego na razie ani na Giełdzie Poznańskiej, ani na Warszawskiej Giełdzie Towarowej, to jest opcji na towarowy kontrakt futures i dokonanie jej wyceny w warunkach polskiego rynku towarowego.

Towarowe kontrakty futures

Największe światowe giełdy to takie, które przeprowadziły w ciągu roku co najmniej milion transakcji handlowych. Jest ich około pięćdziesięciu, a najczęściej wymienia się: Chicago Board of Trade (CBOT), Chicago Mercantile Exchange (CME), Chicago Board of Options Exchange (CBOE), Bolsa de Mercadorias & Futuros w Sao Paulo, London International Financial Futures & Options Exchange (LIFFE), New York Mercantile Exchange (NYMEX), Marche a Terme International de France (MATIF) w Paryżu, Deutsche Terminborse (DTB) we Frankfurcie, London Metal Exchange (LME), Swiss Options and Financial Futures Exchange (SOFFEX) w Zurichu, Tokyo International Financial Futures Exchange (TIFFE), Tokyo Commodity Exchange (TOCOM) oraz OM Stockholm AB.

Handel na tych giełdach zdominowały różnego typu kontrakty terminowe. Ich przedmiotem są także produkty rolne i spożywcze, począwszy od tusz wieprzowych i wołowych, pasz, zboża, kukurydzy i soi, poprzez masło i jaja, po kawę czy cukier. Przy czym rynek futures jest w dużym stopniu wykorzystywany do asekuracji (hedgingu) zabezpieczającej przed ryzykiem i jednocześnie pozwalającej nim zarządzać. Polega ona na sprzedaży lub zakupie kontraktów terminowych dla zrekompensowania ryzyka zmian cen na rynkach gotówkowych. Uczestnicy rynku w celu stworzenia dla siebie ochrony cenowej przyjmują na rynku terminowym pozycję o wartości równej tej, którą otworzyli na rynku gotówkowym. Mechanizm przenoszenia ryzyka spowodował, że kontrakty terminowe stały się niezbędne dla firm i instytucji finansowych na całym świecie, a stosowanie hedgingu ma na celu ochronę zysków, stabilizację przepływu gotówki, ustalenie cen kupna lub sprzedaży określonych walorów, ustalenie równowagi między aktywami danego podmiotu i płynnością rynku, redukowanie kosztów transakcji, obniżanie kosztów magazynowania oraz minimalizowanie potrzeb kapitałowych do utrzymania inwestycji.

Do grupy asekurujących zalicza się: rolników i innych producentów poszukujących ochrony, na przykład przed spadkiem cen zmagazynowanego zboża; właścicieli magazynów poszukujących ochrony przed zmianą cen pomiędzy okresem, kiedy kupują od rolnika zboże lub kontrakt na nie, i okresem, gdy zakup jest rozliczany lub ostatecznie odbędzie się handel zbożem; przetwórców poszukujących ochrony przed wzrostem kosztów surowca; eksporterów poszukujących ochrony przed wzrostem kosztów towarów jeszcze nie nabytych, ale już zakontraktowanych do przyszłej dostawy importerom; hodowców zwierząt i odbiorców pasz poszukujących ochrony przed spadkiem cen zwierząt i wzrostem cen pasz.

Cena kontraktu futures, inaczej cena terminowa, to ustalona w dniu dzisiejszym cena realizacji transakcji na rynku gotówkowym. Mechanizm funkcjonowania rynku terminowego powoduje, że cena ta kształtuje się na rynku jako pochodna bieżącej ceny towaru na rynku gotówkowym w momencie zawarcia transakcji terminowej oraz kosztów przechowania towaru od daty zawarcia kontraktu terminowego do daty jego wykonania. Zmiany cen terminowych na prawidłowo funkcjonującym rynku są wprost proporcjonalne do zmian cen na rynku gotówkowym, co umożliwia wykorzystywanie kontraktów terminowych do zabezpieczania przed ryzykiem zmiany ceny, gdyż wzrost ceny towaru na rynku gotówkowym jest rekompensowany wzrostem ceny kontraktu, a spadek ceny na rynku gotówkowym – spadkiem ceny kontraktu terminowego.

Kontrakty terminowe mają wiele zalet, z których najważniejszą jest bardzo niski koszt uzyskania za ich pomocą zabezpieczenia cenowego, ograniczający się do bardzo niskich prowizji maklerskich oraz – w przypadku kontraktów futures – depozytu zabezpieczającego. Jego wysokość zazwyczaj nie przekracza kilku procent wartości towaru, co więcej, może on być oprocentowany, a po zamknięciu pozycji na rynku futures, pomniejszony o ewentualną stratę lub powiększony o zysk z transakcji, podlega zwrotowi.

Koncentracja obrotów na giełdzie, standaryzacja jednostki transakcyjnej oraz funkcjonowanie izby rozliczeniowej ułatwiają zamykanie wcześniej otwartych pozycji, co umożliwia wywiązanie się ze zobowiązań umownych bez dokonywania odbioru fizycznej dostawy towaru. Jest to szczególnie ważne, kiedy trudno z góry dokładnie określić ilość towaru, którego cenę zakupu lub sprzedaży się zabezpiecza [6].

W Polsce obrót kontraktami terminowymi futures na produkty rolne prowadzi Giełda Poznańska. W 1998 roku były to kontrakty futures na pszenicę konsumpcyjną, od kwietnia 1999 również kontrakty na pszenicę paszową, a od listopada 1999 – na żywiec wieprzowy. Na Warszawskiej Giełdzie Towarowej kontrakty futures na pszenicę konsumpcyjną i paszową oraz żywiec wieprzowy wprowadzono dopiero w 1999 roku.

Opcje towarowe

Najbardziej rozwijającą się dziedziną handlu na światowych giełdach jest handel opcjami. Główna różnica między kontraktem terminowym a opcją polega na tym, że zawierający kontrakt jest zobligowany do jego realizacji bez względu na to, czy przyniesie mu to zysk, czy straty. Natomiast posiadacz opcji ma prawo, ale nie obowiązek, wykonania opcji, za co płaci tzw. premię wystawcy opcji, który ma obowiązek jej wykonania na żądanie kupującego. Opcje, również towarowe, są formą ekonomicznego zabezpieczenia się przed ryzykiem niekorzystnej zmiany ceny instrumentu bazowego (towaru) w transakcjach przyszłych.

Można wyróżnić opcje:

- kupna (call): nabywca opcji uzyskuje prawo zakupu określonej ilości towaru po cenie określonej z góry w okresie do wyznaczonego terminu wygaśnięcia opcji,
- sprzedaży (put): w tym wypadku kupujący opcję nabywa prawo sprzedaży określonej ilości towaru po ustalonej cenie w okresie ważności opcji,
- opcje podwójne (double options): nabywca opcji uzyskuje prawo do zdeklarowania do określonego terminu, czy dokona sprzedaży czy zakupu danej ilości towaru po określonej, w momencie zawarcia transakcji, cenie [4].

Znaczącą część opcji towarowych stanowią opcje na towarowe kontrakty terminowe. Kupujący opcję na kontrakty terminowe może zlikwidować swoją pozycję wykonując opcję lub, podobnie jak w przypadku samego kontraktu terminowego, zamknąć poprzez przeprowadzenie transakcji offsetowej (tzn. zawarcie transakcji odwrotnej do danej, ale o takiej samej wartości).

W ostatnich latach obserwuje się wzrost znaczenia zarządzania pieniędzmi na rynkach terminowych. Wiele instytucji i inwestorów indywidualnych na całym świecie włączyło zarządzanie pozycjami terminowymi do swoich portfeli. Pozycje te występują często w formie funduszy towarowych. Prywatny lub publiczny fundusz towarowy działa jak wzajemny fundusz akcyjny dla tych inwestorów, którzy mają ograniczone środki i czas na inwestycje profesjonalnie zarządzane. Fundusze towarowe są bardzo atrakcyjne, ponieważ dają więcej możliwości lewarowania (zaangażowania niewielkiego kapitału pozwalającego kontrolować poważne transakcje) niż wzajemne fundusze akcyjne. Dodatkowo, niektóre fundusze towarowe, nazywane gwarantowanymi, zapewniają zwrot podstawowej sumy kapitału po okresie inwestycji, a inne zapewniają wypłatę odsetek od zainwestowanego kapitału [1].

W Polsce opcjami towarowymi można na razie handlować na Giełdzie Poznańskiej, na której pierwsza emisja opcji kupna na półtusze wieprzowe miała miejsce 6 czerwca 1995 roku, a w późniejszym okresie rozszerzono przedmiot kontraktu opcyjnego na inne produkty rolne. Na Warszawskiej Giełdzie Towarowej po raz pierwszy można było nabyć towarowe opcje kupna w 1997 roku, a opiewały one na pszenicę konsumpcyjną i paszową, żyto, kukurydzę oraz półtusze wieprzowe i ćwierćtusze wołowe. Jak do tej pory, właściwie jedynym emitentem opcji jest Agencja Rynku Rolnego (ARR), a ustalanie wartości premii następowało w wyniku licytacji po ustaleniu wywoławczej wartości. Ponadto, na obu giełdach prowadzony był wtórny obrót opcjami towarowymi.

Niewątpliwie na cenę towarów w dużym stopniu wpływa ryzyko związane z prowadzeniem biznesu. Ryzyko to można znacząco ograniczać poprzez efektywne wykorzystywanie rynków terminowych, co prowadzi jednocześnie do obniżania kosztów prowadzenia biznesu. Z tych zalet mogą korzystać wszyscy, stąd należy oczekiwać rozwoju rynku terminowego w naszym kraju oraz stopniowego wprowadzania coraz bardziej złożonych instrumentów. Wydaje się, że kolejnym krokiem powinno być wprowadzenie opcji na towarowe kontrakty futures, co prawdopodobnie niedługo nastąpi na WGT.

Metody wyceny opcji

Istotnym problemem związanym z wyceną opcji, również towarowych, jest właściwa wycena opcji, czyli ustalenie wysokości premii opcyjnej. Najczęściej wymieniane i opisywane w literaturze przedmiotu są dwie podstawowe metody wyceny opcji: model Blacka-Scholesa oraz model Coxa-Rossa-Rubinsteina. Pierwsza z metod, opisująca zmiany cen instrumentu bazowego w sposób ciągły, może być wykorzystywana do wyceny opcji europejskich (również towarowych) i amerykańskiej opcji kupna akcji nie wypłacającej dywidendy. Model Coxa-Rossa-Rubinsteina (nazywany również dwumianowym), opisujący zmiany cen instrumentu bazowego w sposób skokowy, można stosować zarówno do wyceny opcji europejskich, jak i amerykańskich, a więc jest metodą bardziej wszechstronną. Ze względu na dużą liczbę opracowań opisujących obydwie wspomniane metody nie będziemy szczegółowo omawiać ich w pracy.

Do wyceny opcji na kontrakty futures oprócz drzew dwumianowych stosuje się model Blacka, pozwalający wycenić opcje europejskie, który zakłada logarytmiczno-normalny rozkład ceny terminowej oraz to, że opcje na kontrakty futures można wyceniać podobnie jak opcje na akcje spółek wypłacają-

cych dywidendę w sposób ciągły. Ponadto, wartość europejskiej opcji na kontrakty futures jest identyczna jak wartość odpowiadającej jej europejskiej opcji natychmiastowej, jeśli założymy, że daty wygaśnięcia obu opcji są takie same. Twierdzenie to nie jest prawdziwe w odniesieniu do opcji amerykańskich, gdyż w warunkach rynku normalnego amerykańska opcja kupna kontraktów futures jest warta więcej niż natychmiastowa amerykańska opcja kupna. W warunkach rynku odwróconego mamy do czynienia z sytuacją przeciwną [3].

Formuła Blacka ma następującą postać:

$$C = e^{-rT}\left[FN(d_1) - XN(d_2)\right] \tag{1}$$

C – cena europejskiej opcji kupna kontraktu futures,

$$P = e^{-rT}\left[XN(-d_2) - FN(-d_1)\right] \tag{2}$$

P – cena europejskiej opcji sprzedaży kontraktu futures,

gdzie:

$$d_1 = \frac{\ln\left(\frac{F}{X}\right) + \sigma^2 \cdot \frac{T}{2}}{\sigma\sqrt{T}} \tag{3}$$

$$d_2 = \frac{\ln\left(\frac{F}{X}\right) - \sigma^2 \cdot \frac{T}{2}}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T} \tag{4}$$

σ oznacza zmienność ceny terminowej, F – aktualną cenę terminową, X – cenę wykonania, r – wolną od ryzyka stopę procentową, T – termin wygaśnięcia.

Formuła Blacka nie wymaga, by termin wygaśnięcia opcji był identyczny z terminem dostawy kontraktu futures.

Przykład wyceny opcji na towarowy kontrakt futures

Na świecie eksperymentalny obrót opcjami na kontrakty futures został zapoczątkowany w 1982 r., a od 1987 r. są one stałym elementem rynków terminowych i od tego czasu ich popularność wzrasta. Podstawową przyczyną tego zjawiska jest fakt, że w wielu przypadkach płynność rynku kontraktów futures jest większa niż płynność aktywów pierwotnych i obrót nimi jest prostszy. Również nie ulega wątpliwości, że łatwiej i wygodniej jest przyjąć lub wykonać dostawę kontraktu futures, na przykład na żywiec wieprzowy, niż dostawę samego żywca. Zwykle wykonanie opcji na kontrakt futures nie prowadzi do dostawy instrumentu bazowego, gdyż pozycja futures jest najczęściej zamykana przed nadejściem terminu dostawy i dlatego opcje na kontrakty futures są na ogół rozliczane gotówkowo, co zwiększa ich atrakcyjność (szczególnie dla inwestorów nie posiadających, w razie wykonania opcji, środków wystarczających do zakupu towaru bazowego). Udogodnieniem jest także fakt, że obrót opcjami na kontrakty futures i właściwymi kontraktami futures odbywa się na światowych giełdach w sąsiadujących ringach, co ułatwia dokonywanie transakcji zabezpieczających, arbitrażowych czy spekulacyjnych i w konsekwencji zwiększa efektywność rynku. Ponadto, z opcjami na kontrakty futures często są związane niższe koszty transakcyjne niż koszty opcji natychmiastowych [3].

W Polsce do tej pory nie ma możliwości handlowania opcjami na kontrakty futures i dlatego przedstawiony w pracy przykład wyceny opcji na towarowy kontrakt futures stanowi symulację, wykorzystującą jednak elementy rzeczywistego rynku terminowego. Proponowana opcja opiewa na kontrakt futures, którego przedmiotem jest pszenica konsumpcyjna. Kontrakty takie są notowane na Giełdzie Poznańskiej, dlatego do obliczeń zostaną wykorzystane dane zamieszczane w *Towarniku* – miesięcznym biuletynie informacyjnym GP (nr 39/02-2000 i 43/06-2000). Do przeprowadzenia wyceny opcji potrzebne są pewne informacje wyjściowe: rozpatrywana opcja została wystawiona 05.01.2000 roku, a termin realizacji kontraktu futures przypada w maju 2000 roku, zatem okres ważności opcji T = 120 dni. Cena wykonania X została ustalona w wysokości 530 zł/t, zaś F, czyli aktualna cena terminowa, kształtuje się na poziomie 520 zł/t. Założona stopa wolna od ryzyka r = 14,4%, a zmienność ceny terminowej $\sigma = 23\%$.

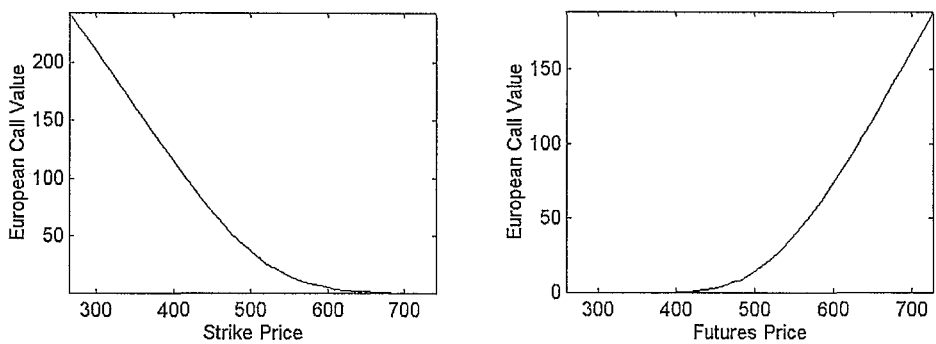
Zgodnie z omówioną w pracy metodyką wyceny opcji na kontrakty futures, wykorzystując model Blacka, można obliczyć wartość europejskiej opcji kupna (C) oraz sprzedaży (P), do czego został wykorzystany pakiet *The Black-*

-*Scholes And Beyond Interactive Toolkit*. Uzyskane wyniki są następujące: C = 21,82 i P = 31,35. Jeśli zastosować metodę drzew dwumianowych, która jest dostępna w wymienionym pakiecie, można wycenić również opcje amerykańskie. Wyniki wyceny modelem dwumianowym są następujące:

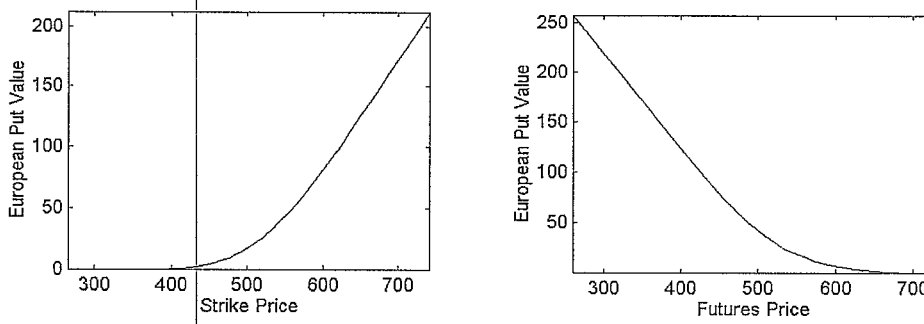
- wartość opcji kupna: europejskiej C = 21,86 i amerykańskiej c = 22,07;
- wartość opcji sprzedaży: europejskiej P = 31,40 i amerykańskiej p = 31,77.

Porównanie parametrów europejskiej i amerykańskiej opcji kupna daje zbliżone wartości Delta, Gamma, Theta i Rho. Natomiast występuje znaczna różnica wartości parametru Vega, który dla opcji europejskiej wynosi 1,14, a dla amerykańskiej 22,44. Dla opcji sprzedaży Vega wynosi odpowiednio: 1,14 i 38,45. Wysoka wartość tego parametru świadczy o dużej zależności między wartością opcji amerykańskiej i zmiennością ceny terminowej.

Posiadacz europejskiej opcji kupna mógł w maju 2000 roku dokonać realizacji kontraktu futures, który gwarantował mu zakup pszenicy konsumpcyjnej po cenie 530 zł/t, podczas gdy na rynku gotówkowym w tym czasie średnia cena kształtowała się na poziomie 638,83 zł/t. Dla niego była to transakcja opłacalna, czego nie można powiedzieć o posiadaczu opcji sprzedaży, który na wykonaniu tej opcji straciłby. Rysunki 1 i 2 przedstawiają zależność między wartością europejskiej opcji kupna i sprzedaży kontraktu futures a ceną wykonania i ceną terminową.



Rysunek 1
Zależność pomiędzy ceną wykonania oraz ceną terminową a wartością europejskiej opcji kupna kontraktu futures



Rysunek 2
Zależność pomiędzy ceną wykonania oraz ceną terminową a wartością europejskiej opcji sprzedaży kontraktu futures

Podsumowanie

We współczesnej gospodarce giełdy niewątpliwie odgrywają ogromną rolę i przynoszą zarówno korzyści bezpośrednie, jak i pośrednie. Przede wszystkim zwiększają tzw. płynność handlowania towarami, a im większa płynność, tym więcej możliwości rozszerzania rynku i skuteczniejsze możliwości finansowania powstających niekorzystnych różnic cenowych na rynkach terminowych. Wprowadzenie transakcji terminowych spowodowało jedno z najefektywniejszych przeobrażeń w handlu towarami. Uwzględniają one nawet takie zjawiska, jak sezonowe wahania spożycia. Ponadto, towar sprzedawany na giełdach jest standaryzowany i odpowiedniej jakości. Dzięki nim powstał więc system narzucający zasady nie tylko bezpośrednim uczestnikom gry giełdowej, ale wpływający także na gospodarki poszczególnych krajów, a pośrednio też na gospodarkę światową.

Giełdy, na których zawierane są kontrakty terminowe i opcyjne, przynoszą wiele korzyści ekonomicznych, związanych przede wszystkim z ograniczaniem ryzyka finansowego, możliwościami inwestycji lewarowych i zwiększeniem płynności. Poza tym, poprzez przyjęcie odpowiednich reguł handlu giełdy zabezpieczają rynek przed powstawaniem karteli, które mogłyby narzucać ceny, a przez to zdobyć pozycję monopolisty. Co więcej, z rozwojem nowoczesnych form handlu towarowego wiążą się korzyści, które rzadko jesteśmy skłonni im przypisywać, chociażby rozwój nowoczesnych sposobów przekazywania informacji czy normy prawne wykorzystywane chętnie przez inne podmioty, na przykład firmy.

Największymi instytucjami współpracującymi z giełdami są banki. Aczkolwiek pracują one oddzielnie, to jednak tworzą układ wzajemnie się uzupełniający. Odnosi się to szczególnie do ostatnich dwudziestu lat, kiedy to giełdy przejęły sposób ustalania wzajemnych kursów głównych walut świata. Giełdy napędzają koniunkturę, banki posiadają kapitał i udzielają kredytów. Aby jednak taki system mógł funkcjonować, potrzebne są określone warunki.

Specjaliści przewidują dwie możliwe koncepcje dotyczące przyszłości giełd towarowych. Pierwsza mówi o powstawaniu nowych giełd, głównie w Azji Południowej i Europie Wschodniej, które chociaż nigdy nie będą konkurować z największymi takimi instytucjami, mogą jednak odegrać ważną rolę w tworzeniu globalnego systemu handlowego. Druga koncepcja, prawdopodobna w obliczu ostatnich głośnych fuzji, przewiduje łączenie interesów przez największe giełdy świata. Ponadto, wprowadza się coraz to nowe kontrakty terminowe, a dynamiczny rozwój techniki sprzyja tworzeniu nowych produktów.

Największe polskie giełdy towarowe – Giełda Poznańska i Warszawska Giełda Towarowa – funkcjonują już na zasadach zbliżonych do tych, według których działają giełdy zachodnie, chociaż wciąż przeważają na nich transakcje rzeczywiste, gotówkowe. Jednocześnie jest to rynek ułomny, gdyż jak na razie dostępne są tylko opcje kupna, których jedynym emitentem jest ARR. Nie pozwala to zatem na wykorzystanie chociażby opcyjnych strategii inwestycyjnych, mających szerokie zastosowanie na światowych giełdach. Można się spodziewać rozwoju rynku instrumentów pochodnych w Polsce, należy jednak podjąć działania informacyjne, gdyż ciągle poziom wiedzy potencjalnych inwestorów na ich temat jest niewystarczający.

Literatura

1. BLIŹNIAK D., GONTARSKI L., 1996: *Giełda towarowa.* Fundacja na Rzecz Giełdy Zbożowo-Paszowej, Warszawa.
2. CHRISS N., 1997: *The Black-Scholes And Beyond Interactive Toolkit.* McGraw-Hill Company, New York.
3. HULL J., 1998: *Kontrakty terminowe i opcje.* WIG-Press, Warszawa.
4. JERZAK M., 1998: *Giełda towarowa na rynku rolnym.* Fundacja na Rzecz Giełdy Zbożowo-Paszowej, Warszawa.
5. KRAWIEC B., WROŃSKI S., 2000: Instrumenty pochodne na giełdach towarowych w Polsce. *Mat. V Zachodniopomorskiego Forum Finanse 2000.* Wydawnictwo Szczecin-Expo.

6. REMBISZ W., 1999: Rynek terminowy jako forma ograniczania ryzyka na rynku rolnym. *Mat. IV Zachodniopomorskiego Forum Finanse '99.* Wydawnictwo Szczecin-Expo.

Commodity futures option pricing in conditions of the Polish market

Abstract

Since 1970s the importance of financial derivatives, designed to help manage risk, has grown significantly. These instruments are forwards, futures, swaps and options. However, many models and methods created for the needs of financial engineering may be successfully employed on commodity markets.

In the paper, there are described commodity futures and options available on world and Polish commodity exchanges. There are also presented futures options pricing methods, e.g. the Black model used for European options pricing and the Cox-Ross-Rubinstein method employed for European and American options evaluation.

Then, an option on futures on consumption wheat is created. To this aim, data from the Poznań Exchange, where commodity future contracts are traded, has been used.

The proposed option is priced by the use of an computer program The Black-Scholes And Beyond Interactive Toolkit which gives a possibility of employing both described methods.

Ewa Marta Syczewska
Instytut Ekonometrii SGH W Warszawie

Testy kointegracji dla szeregów I(2), I(1) – Testy Hansena

Testy Hansena

Przedmiotem niniejszych rozważań jest zastosowanie testów Hansena do badania stabilności modelu, w którym zmienna objaśniana Y jest zintegrowana stopnia 1, natomiast dwie zmienne objaśniające X1 i X2 są zintegrowane stopnia I(2), zarazem są skointegrowane do stopnia I(1). Ponadto, zmienne Y, X1 i X2 są skointegrowane, więc składnik losowy takiego równania jest stacjonarny (por. np. Engle i Granger [1987] lub Charemza i Deadman [1997]).

Testy Hansena – Lc, MeanF oraz SupF – związane są z semiparametryczną metodą estymacji relacji kointegrujących, opracowaną przez Phillipsa i Hansena (por. Phillips i Hansen [1990]) i uogólnioną następnie przez Hansena [1992a] na przypadek modelu o zmiennych parametrach. Omówienie tej metody oraz własności testów Hansena, w tym tablice wartości krytycznych, dla przypadku zmiennych objaśniających zintegrowanych stopnia 1, oraz argumenty za stosowaniem tych metod podałam w mojej pracy (Syczewska [1999]).

Hansen [1992] pisze, że wszystkie trzy testy, tzn. SupF, MeanF oraz Lc, są testami tej samej hipotezy zerowej, zakładającej stabilność (brak zmian) parametrów strukturalnych szacowanego modelu. Testy te różnią się natomiast postacią hipotezy alternatywnej, zakłada ona bowiem albo jednorazową zmianę strukturalną w nieznanym momencie, albo stopniową zmianę parametrów (parametry generowane są jako martyngały). Najłatwiejsza do obliczenia jest statystyka Lc. Nie jest ona obarczona ograniczeniem, jakiemu podlegają dwie pozostałe statystyki, dla których z przyczyn analitycznych (co wykazał Andrews [1990]) trzeba zakładać, że ewentualna zmiana mogła wystąpić tylko w części próby odpowiadającej proporcji z przedziału $\langle 0{,}15\ 0{,}85\rangle$, a nie w dowolnym momencie (tzn. nie dla przedziału $\langle 0\ 1\rangle$). Testy Hansena mogą być wykorzystywane jako testy hipotezy zerowej, oznaczającej występowanie relacji kointegrującej.

Statystyki testów Hansena mają złożoną postać. Rozkłady asymptotyczne dla tych statystyk, podane przez Hansena [1991], [1992] dla zmiennych I(1), są skomplikowanymi funkcjami procesów Wienera. Zadanie zbadania metodami analitycznymi własności testów dla przypadku, gdy zmienne objaśniające są zintegrowane stopnia wyższego niż 1, jest znacznie bardziej skomplikowane. Tymczasem przypadek, stanowiącego potencjalną relację kointegrującą, modelu o zmiennych objaśniających zintegrowanych stopnia 2 nie jest rzadki. Przy badaniu relacji parytetu siły nabywczej zmienne odpowiadające poziomowi cen w kraju i za granicą lub oznaczające nominalny kurs wymiany bywają zmiennymi I(2) (por. np. Haldrup [1994]). Potrzebne są więc odpowiednie tablice wartości krytycznych testów Hansena. Otrzymałam je wykorzystując metody Monte Carlo. Obliczenia wykonane są przy użyciu pakietu COINT w programie GAUSS wersja 3.2.14 dla systemu DOS oraz – w przeważającej części – w programie GAUSS wersja 3.2.52 dla Linuxa, ze względu na stabilność tego systemu operacyjnego.

Procedura FM(y,x,d,l), oprogramowana w GAUSSIE w pakiecie COINT, służy do obliczania estymatora FM (Fully Modified) Phillipsa-Hansena oraz statystyk testów Hansena dla regresji skointegrowanej, wykorzystuje metodę najmniejszych kwadratów dla regresji na pierwszym etapie obliczeń. Składnia polecenia jest następująca (por. Ouliaris i Phillips [1994]):

{beta, vc, stderr, sigma, tstats, rss, resid, tests} = fm(y,x,d,k),

gdzie argumentami wejściowymi są:
y – wektor kolumnowy obserwacji zmiennej objaśnianej,
x – macierz obserwacji zmiennych objaśniających,
d – określa część deterministyczną równania regresji,
k – określa liczbę składników autokowariancji wykorzystywanych do obliczania spektrum przy częstotliwości zero.
Wynikiem obliczeń są następujące wielkości:
beta – wektor kolumnowy zawierający oceny parametrów badanej relacji,
vc – macierz wariancji ocen parametrów,
stderr – wektor odchyleń standardowych ocen parametrów,
sigma – błąd standardowy reszt,
tstats – statystyki *t* dla ocen parametrów,
rss – suma kwadratów reszt,
resid – wektor kolumnowy reszt,
stests – wektor kolumnowy wymiaru 3×1, którego kolejnymi elementami są wartości statystyk testów Hansena, służących testowaniu hipotezy zerowej,

oznaczającej, że wektor kointegrujący jest stabilny w badanym okresie. Kolejność testów jest następująca: Lc, Mean F i SupF. Metoda estymacji jest opracowana w pakiecie COINT według pracy Phillipsa i Hansena [1990], natomiast testy – według pracy Hansena [1991].

W celu wykorzystania metody estymacji i testów Hansena do badania stabilności relacji kointegrującej o zmiennych objaśniających I(2), typu modelu opisującego kursy walutowe, należało zbadać i opracować tablice wartości krytycznych testów Hansena przy następujących założeniach:

- w równaniu regresji występują dwie zmienne objaśniające;
- obie są zmiennymi zintegrowanymi stopnia 2;
- są skointegrowane do wartości I(1);
- natomiast zmienna objaśniana jest zintegrowana stopnia I(1).

Taką wersję modelu w dalszych rozważaniach nazywam wariantem 2, natomiast wariant 1 jest to model również jednorównaniowy z dwiema zmiennymi objaśniającymi zintegrowanymi stopnia 1. Oba stanowią relacje kointegrujące o stałych parametrach (czyli prawdziwa jest hipoteza zerowa o stałości parametrów modelu).

Porównanie rozkładów testów dla zmiennych I(1) i dla zmiennych I(2)

Interesujące jest sprawdzenie, czy wartości krytyczne rozkładu statystyk testów Hansena są zbliżone, czy też znacznie się różnią w przypadku, gdy wszystkie zmienne objaśniające są zmiennymi I(1) i w przypadku, gdy dwie zmienne objaśniające są I(2) skointegrowane do I(1). Porównałam w tym celu wyznaczone na podstawie 100 000 replikacji empiryczne rozkłady wygenerowane przy tej samej wartości parametrów regresji (czyli takich samych współczynnikach przy X1 i X2) dla wariantów 1 i 2, a zatem dla modeli różniących się jedynie założeniami co do zintegrowania zmiennych objaśniających. Do sprawdzenia wniosku o zróżnicowaniu rozkładów wykorzystałam nieparametryczny test rang w wersji omówionej przez Pawłowskiego [1980, str. 240–243]. Wektory kolumnowe obliczonych wartości testów łączymy w macierz A, konstruujemy macierz rang R o tym samym wymiarze, przy czym rangi są równe 1 lub 2 w zależności od tego, czy pierwszy element macierzy A jest większy od drugiego, czy przeciwnie; dla jednakowych wartości elementów przyjmuje się rangę równą średniej arytmetycznej odpowiednich wartości.

W ogólnym przypadku rangi przyjmują wartości od 1 do r. Średnia elementów macierzy rang jest równa:

$$\bar{t} = \frac{(1+r)n}{2},$$

gdzie r jest najwyższą rangą, zaś n odpowiada liczebności próby.
W naszym przypadku $r = 2$, $n = 100\,000$, zatem $\bar{t} = 150\,000$. Jako sprawdzian hipotezy zerowej przyjmuje się statystykę:

$$Q_T^2 = \sum_i \frac{(t_i - \bar{t})^2}{\bar{t}},$$

gdzie t_i oznacza w tym przypadku sumę rang w kolumnie. Pawłowski [1980, str. 241] podaje, że przy dużych n i gdy hipoteza zerowa o tym, że rozkłady są identyczne, jest prawdziwa, statystyka ta ma w przybliżeniu rozkład chi-kwadrat o $r - 1$ stopniach swobody.

Statystyka ta została obliczona dla każdego z testów Lc, MeanF i SupF, na podstawie wszystkich wartości wygenerowanych dla modelu ze zmiennymi I(1) oraz dla modelu ze zmiennymi I(2). Obliczenia powtórzono zarówno dla wartości pierwotnych, jak i dla posortowanych według wielkości. Dodatkowo zliczono liczbę dodatnich różnic pomiędzy wartościami testów otrzymanych dla tych dwu wariantów. Statystyki i współczynniki asymetrii i spłaszczenia dla testów Lc, MeanF i SupF przy pierwszym i drugim wariancie zintegrowania zmiennych objaśniających podane są w tabeli 1.

Tabela 1
Porównanie testów Hansena dla dwu wariantów zintegrowania zmiennych objaśniających

Test i wariant	Lc, 1	Lc, 2	MeanF, 1	MeanF, 2	SupF, 1	SupF, 2
Średnie	0,6037	4,7642	9,0420	0,2959	3,4237	9,4535
Minima	0,0830	0,6980	2,0120	0,0629	0,7138	2,2916
Maksima	2,5840	15,2134	155,9002	2,7575	23,4246	277,0623
Odchylenie standardowe	0,3144	1,8045	2,8953	0,1650	1,3925	4,1746
Współczynnik asymetrii	1,0735	1,9889	0,6088	1,4268	0,6088	1,4268
Współczynnik spłaszczenia	1,1811	6,4681	0,0590	4,1076	0,0590	4,1076

Test	Wartość statystyki Q^2	Liczba dodatnich różnic
Lc	157638,9	83859
MeanF	153593,1	73222
SupF	149998,0	50000

Źródło: Obliczenia własne.

Otrzymane wyniki wskazują jednoznacznie na to, że hipotezę zerową (oznaczającą jednakowy rozkład niezależnie od stopnia zintegrowania zmiennych) należy odrzucić dla wszystkich trzech testów Hansena.Wniosek ten jest zgodny ze stwierdzeniem Hansena [1992], odnoszącym się wprawdzie do modeli, w których zmienne objaśniające były bądź I(1), bądź zmiennymi deterministycznymi. Hansen podkreśla, że rozkłady statystyk zależą od stopnia zintegrowania i od kowariancji zmiennych objaśniających.

Tabele wartości krytycznych dla zmiennych objaśniających I(2)

Założenia przyjęte przy generowaniu podanych tabel wartości krytycznych są następujące:

- przyjęte wartości liczebności próby są równe {25, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500};
- dla każdej z tych liczebności, przy założeniu stabilności parametrów modelu, zostało wygenerowane po 20 000 replikacji procesu generującego dane;
- dla każdej replikacji oszacowano model oraz obliczono wartości statystyk testów Hansena;
- na tej podstawie obliczono empiryczne percentyle rozkładów statystyk przedstawione w tabeli 2.

Tabela 2
Wartości krytyczne testów Lc, MeanF i SupF dla zmiennych I(2), na podstawie 20 000 replikacji

Test Lc

L. obs.	99%	97,5%	95%	90%	10%	5%	2,5%	1%
25	1,2416	1,0032	0,8309	0,6776	0,2050	0,1807	0,1621	0,1441
50	0,8714	0,7268	0,6137	0,5051	0,1441	0,1277	0,1159	0,1039
75	0,8572	0,7011	0,5902	0,4764	0,1250	0,1090	0,0978	0,08698
100	0,8233	0,6742	0,5695	0,4599	0,1168	0,1010	0,09001	0,07937
125	0,8671	0,7023	0,5826	0,4616	0,1108	0,0952	0,08361	0,07414
150	0,8445	0,6844	0,5726	0,4580	0,1082	0,09218	0,08113	0,07117
175	0,8473	0,6933	0,5644	0,4526	0,1067	0,09031	0,07895	0,06877
200	0,8356	0,6815	0,5679	0,4483	0,1052	0,08947	0,07872	0,06845
225	0,8348	0,6860	0,5703	0,4579	0,1047	0,08861	0,07678	0,06644

Test Lc

L. obs.	99%	97,5%	95%	90%	10%	5%	2,5%	1%
275	0,8468	0,6755	0,5605	0,4459	0,1026	0,08674	0,07583	0,06558
250	0,8515	0,6930	0,5665	0,4507	0,1037	0,08760	0,07675	0,06590
300	0,8460	0,6935	0,5720	0,4482	0,1014	0,08557	0,07425	0,06398
325	0,8414	0,6772	0,5577	0,4459	0,1007	0,08591	0,07398	0,06282
350	0,8630	0,6800	0,5566	0,4450	0,1011	0,08436	0,07345	0,06346
375	0,8596	0,6806	0,5657	0,4425	0,09916	0,08345	0,07194	0,06179
400	0,8399	0,6752	0,5495	0,4413	0,09936	0,08412	0,07296	0,06316
425	0,8506	0,6847	0,5602	0,4434	0,09913	0,08371	0,07183	0,06165
450	0,8399	0,6906	0,5595	0,4436	0,09934	0,08308	0,07058	0,06027
475	0,8542	0,6860	0,5651	0,4473	0,09862	0,08305	0,07212	0,06188
500	0,8327	0,6718	0,5558	0,4436	0,09800	0,08153	0,07034	0,05967

cd. tabeli 2

Test MeanF

L. obs.	99%	97,5%	95%	90%	10%	5%	2,5%	1%
25	28,8366	17,0771	12,8377	9,7703	2,9106	2,5706	2,3052	2,0563
50	7,9097	6,8085	6,0259	5,0259	1,9682	1,7554	1,5785	1,3816
75	7,4140	6,3784	5,6235	4,8645	1,6791	1,4543	1,3013	1,1313
100	7,2748	6,2772	5,4978	4,7055	1,5518	1,3208	1,1763	1,0181
150	7,4452	6,3809	5,5397	4,6960	1,4306	1,2223	1,0750	0,9135
175	7,4950	6,3730	5,5994	4,7174	1,4141	1,1981	1,0476	0,8954
200	7,4586	6,3961	5,5776	4,7008	1,3962	1,1876	1,0318	0,8817
225	7,5736	6,4415	5,6089	4,7664	1,3763	1,1731	1,0106	0,8590
250	7,4936	6,4205	5,5582	4,6872	1,3688	1,1571	0,9969	0,8585
275	7,6474	6,4702	5,5671	4,6727	1,3625	1,1553	1,0064	0,8553
300	7,7001	6,5301	5,6297	4,6903	1,3315	1,1173	0,9592	0,8176
325	7,6190	6,4564	5,5737	4,7125	1,3375	1,1361	0,9777	0,8302
350	7,7820	6,5203	5,6130	4,7579	1,3215	1,1050	0,9553	0,8096
375	7,6684	6,4901	5,5783	4,6658	1,3163	1,1018	0,9604	0,8142
400	7,5855	6,5262	5,5985	4,7009	1,3212	1,1069	0,9557	0,8198
425	7,6518	6,5839	5,6839	4,7291	1,3111	1,1034	0,9535	0,8017
450	7,7317	6,5210	5,6449	4,6765	1,3226	1,1081	0,9499	0,8027
475	7,7648	6,6240	5,6839	4,7542	1,3174	1,1050	0,9512	0,8092
500	7,6462	6,4417	5,5720	4,6736	1,2860	1,0778	0,9328	0,7858

Test SupF

L. obs.	99%	97,5%	95%	90%	10%	5%	2,5%	1%
25	257,2601	105,7813	59,6513	34,3281	7,1206	6,2576	5,6100	4,9485
50	22,2578	18,3481	15,7444	13,6039	5,8947	5,2943	4,8131	4,3315
75	16,6868	14,7788	13,3525	12,0141	5,4795	4,8758	4,3782	3,8685
100	15,6458	14,2601	12,9666	11,6702	5,2244	4,6426	4,2026	3,7417
150	15,7594	14,3406	13,1925	11,8229	5,1497	4,5510	4,0997	3,6359
175	16,4408	14,6162	13,2397	11,8731	5,1309	4,4999	4,0331	3,5482
200	16,1397	14,5891	13,3466	11,9194	5,1251	4,4773	4,0296	3,5507
225	16,5227	14,8858	13,5171	12,1025	5,0967	4,4950	4,0026	3,5381
250	16,6707	14,9057	13,5556	12,1127	5,0951	4,5189	4,0381	3,5579
275	16,5939	14,8997	13,5918	12,1269	5,0761	4,4953	4,0288	3,5367
300	16,7930	15,0340	13,6698	12,1992	5,0705	4,4453	3,9803	3,4351

cd. tabeli 2

L. obs.	99%	97,5%	95%	90%	10%	5%	2,5%	1%
325	17,0318	15,2575	13,7833	12,2435	5,0828	4,4783	4,0070	3,4968
350	17,2450	15,3597	13,8872	12,2620	5,0477	4,4512	3,9766	3,4625
375	17,1593	15,1172	13,7321	12,2097	5,0777	4,4603	3,9841	3,4598
400	17,0030	15,3130	13,9101	12,4088	5,0903	4,4384	3,9796	3,4899
425	17,3257	15,4488	14,0210	12,4322	5,0857	4,4847	3,9899	3,4928
450	17,4829	15,4346	13,9120	12,3753	5,0874	4,4580	3,9619	3,4627
475	17,4141	15,7087	14,2167	12,5632	5,1012	4,4753	4,0123	3,5505
500	17,3406	15,5068	14,0420	12,4431	5,0509	4,4490	3,9855	3,4918

Źródło: Obliczenia własne.

Powierzchnie odpowiedzi

MacKinnon [1991], analizując własności testów integracji i kointegracji Dickeya-Fullera oraz Engle'a-Grangera, prócz wyznaczania wartości krytycznych metodą Monte Carlo stosuje następującą metodę wyznaczania powierzchni odpowiedzi. Dla każdego wariantu liczebności próby otrzymuje po 40 egzemplarzy wyników (tzn. powtarza obliczenia percentyli po 40 razy, generując za każdym razem odpowiednią liczbę replikacji). Uzyskuje w ten sposób zmienne odpowiadające liczebności próby oraz wektory wartości percentyli, oznaczone symbolem $C_k(p)$. W celu wyznaczenia asymptotycznych wartości krytycznych testu szacuje regresję liniową zmiennej $C_k(p)$ względem dwu zmiennych: T_k^{-1} oraz T_k^{-2}, postaci:

$$C_k(p) = \alpha_0 + \alpha_1 T_k^{-1} + \alpha_2 T_k^{-2} + \varepsilon_k ,$$

gdzie T_k oznacza wektor kolumnowy, którego kolejne elementy odpowiadają liczebnościom próby w kolejnych powtórzeniach eksperymentu. Niech a_0, a_1 i a_2 oznaczają oceny odpowiednich parametrów (tutaj – otrzymane metodą najmniejszych kwadratów). Ocena wyrazu wolnego tej regresji, czyli a_0, stanowi przybliżenie asymptotycznej wartości krytycznej testu, jest bowiem równa granicy wyrażenia $a_0 + a_1 T^{-1} + a_2 T^{-1}$ przy $T \to \infty$. MacKinnon [1994], [1996] wprowadza w późniejszych pracach bardziej rozbudowane warianty analizy powierzchni odpowiedzi, jednakże, jak podkreśla, są one znacznie bardziej pracochłonne (obliczenia przeprowadzone przez MacKinnona na wielu komputerach łącznie zajęły około 2 lat).

Dla przedstawionych tu rozważań przyjęłam mniejszą liczbę powtórzeń eksperymentów i mniejszą – replikacji. Mianowicie każdy eksperyment dla liczebności zbioru obserwacji równej 25, 50, 75, 100, 125, 150, 200 i 300 został powtórzony po 20 razy. (Przypomnijmy, że eksperyment polegał na obliczaniu percentyli rozkładu badanego testu na podstawie 20 000 replikacji.) Otrzymane wyniki przedstawione są w tabeli 3. Pierwsza liczba, wyróżniona pogrubioną czcionką, stanowi ocenę wyrazu wolnego w równaniu regresji dla danej wartości krytycznej, jest zarazem oceną asymptotycznej wartości krytycznej danego testu. W nawiasach przy wszystkich ocenach parametrów podane są błędy ocen. Dopasowanie równań należy ocenić jako dobre, na co wskazują wartości współczynników determinacji.

Tabela 3
Powierzchnie odpowiedzi na podstawie eksperymentów powtórzonych po 20 razy, dla 20 000 replikacji, przy liczebnościach próby T = {25, 50, 75, 100, 125, 150, 200, 300}

Percentyl	99%	95%	90%
	Test Lc		
Stała	**2,5661** (0,0571)	**1,3925** (0,0251)	**1,0654** (0,01607)
a_1	–16,7448 (0,4809)	–8,4335 (0,2112)	–6,3837 (0,1353)
a_2	40,3917 (0,9833)	21,4335 (0,4319)	16,5721 (0,2766)
R^2	0,977	0,989	0,994
	Test MeanF		
Stała	**107,656** (3,6444)	**37,574** (1,002)	**24,718** (0,583)
a_1	–958,925 (30,682)	–309,372 (8,438)	–195,498 (4,907)
a_2	2265,286 (62,736)	739,325 (17,253)	473,372 (10,033)
R^2	0,964	0,977	0,983
	Test SupF		
Stała	**1129,53** (42,35)	**219,18** (6,573)	**111,62** (2,82)
a_1	–10690,53 (356,55)	–1981,53 (55,34)	–961,58 (23,73)
a_2	25350,67 (729,03)	4712,19 (113,15)	2292,18 (48,52)
R^2	0,963	0,974	0,980

Źródło: Obliczenia własne.

Wygładzone wartości krytyczne na podstawie 100 000 replikacji

Dokładność wyników można poprawić dzięki zwiększeniu liczby replikacji. (Wartości krytyczne wyznaczone na podstawie 100 000 replikacji są podane w tabeli 5.) Można jednak dodatkowo zastosować wygładzanie empirycznych percentyli statystyk testów, wykorzystane przez Charemzę i Deadmana (1997), a uwzględniające oceny wariancji obliczonej przez nich według wzoru Davida-Johnsona (por. David i Johnson [1954]).

W przeciwieństwie do wzoru zastosowanego przez Charemzę i Deadmana, czyli specyfikacji z odwrotnościami logarytmów liczebności próby, podane tu wyniki obliczone są dla regresji, w której zmienne objaśniające to odwrotność liczebności próby oraz odwrotność jej kwadratu. Specyfikacja ta jest podobna do tej, jaką zastosowałam do obliczania asymptotycznych wartości krytycznych według metody MacKinnona. Wygładzone wartości krytyczne podane są w tabeli 4.

Podsumowanie i wnioski

Przedstawione tu tabele wartości krytycznych dla testów Hansena stanowią narzędzie potrzebne do przeprowadzenia testów relacji kointegrującej o zmiennych objaśniających zintegrowanych I(2) i skointegrowanych do poziomu I(1), takiej jak na przykład przy badaniu parytetu siły nabywczej. Jednocześnie wyniki te są ilustracją kilku sposobów wyznaczania wartości krytycznych, w tym również asymptotycznych, metodami obliczeniowymi zapewniającymi dobrą dokładność, umożliwiającą wnioskowanie z należytą pewnością. Kontynuacja badań będzie polegała na opracowaniu obszerniejszych tablic wartości krytycznych oraz na ich praktycznym zastosowaniu do modelowania makroekonomicznego.

Tabela 4
Wygładzone wartości krytyczne dla testu Lc

L. obs.	Lc, 99% dolna; górna	Lc, 97,5% dolna; górna	Lc, 95% dolna; górna	Lc, 90% dolna; górna	Lc,10%; dolna; górna	Lc, 5% dolna; górna
25	1,235; 1,322	0,987; 1,048	0,824; 0,862	0,672; 0,700	0,193; 0,215	0,165; 0,193
50	0,856; 0,944	0,711; 0,772	0,604; 0,642	0,496; 0,524	0,132; 0,154	0,111; 0,139
75	0,808; 0,896	0,669; 0,731	0,567; 0,605	0,462; 0,490	0,114; 0,137	0,094; 0,123
100	0,799; 0,886	0,658; 0,720	0,555; 0,593	0,449; 0,478	0,106; 0,128	0,087; 0,115
125	0,798; 0,885	0,655; 0,716	0,550; 0,588	0,443; 0,472	0,101; 0,123	0,082; 0,110
150	0,799; 0,887	0,653; 0,715	0,548; 0,586	0,440; 0,469	0,097; 0,120	0,079; 0,107
200	0,803; 0,890	0,653; 0,715	0,546; 0,584	0,437; 0,465	0,094; 0,116	0,075; 0,103
300	0,809; 0,897	0,655; 0,717	0,545; 0,583	0,434; 0,463	0,090; 0,112	0,071; 0,099
400	0,813; 0,901	0,656; 0,718	0,545; 0,583	0,433; 0,461	0,088; 0,111	0,069; 0,098
500	0,816; 0,904	0,657; 0,719	0,545; 0,583	0,432; 0,461	0,087; 0,109	0,068; 0,097
600	0,818; 0,905	0,658; 0,720	0,545; 0,583	0,432; 0,460	0,087; 0,109	0,068; 0,096

Źródło: Obliczenia własne.

Tabela 5
Wartości krytyczne obliczone na podstawie 100 000 replikacji

Test Lc

L. obs.	99%	97,5%	95%	90%	10%	5%	2,5%	1%
25	1,2813	1,0195	0,8443	0,6866	0,2044	0,1793	0,1611	0,1439
50	0,8840	0,7278	0,6187	0,5075	0,1433	0,1264	0,1147	0,1026
75	0,8651	0,7101	0,5893	0,4771	0,1251	0,1084	0,0973	0,0867
100	0,8471	0,6954	0,5775	0,4648	0,1171	0,1008	0,0896	0,0789
125	0,8448	0,6832	0,5675	0,4568	0,1120	0,0962	0,0849	0,0744
150	0,8482	0,6893	0,5699	0,4565	0,1091	0,0930	0,0818	0,0713
200	0,8653	0,6845	0,5654	0,4523	0,1056	0,0894	0,0784	0,0677
300	0,8511	0,6957	0,5626	0,4464	0,1015	0,0854	0,0741	0,0641
400	0,8507	0,6826	0,5614	0,4473	0,0995	0,08385	0,07278	0,0622
500	0,8539	0,6863	0,5634	0,4449	0,0986	0,08271	0,07161	0,06141
600	0,8543	0,6875	0,5652	0,4471	0,0982	0,08223	0,0718	0,06137

cd. tabeli 5

Test MeanF

L. obs.	99%	97,5%	95%	90%	10%	5%	2,5%	1%
25	28,8062	17,7096	13,0172	9,8124	2,9130	2,5703	2,3101	2,0528
50	7,9134	6,8359	6,0328	5,2380	1,9493	1,7215	1,5482	1,3721
75	7,3116	6,3715	5,6400	4,8527	1,6695	1,4535	1,2939	1,1277
100	7,3263	6,3443	5,5669	4,7681	1,5429	1,3276	1,1665	1,0126
125	7,3515	6,3458	5,5485	4,7043	1,4756	1,2670	1,1093	0,9547
150	7,4098	6,3804	5,5672	4,7154	1,4430	1,2325	1,0711	0,9232
200	7,5623	6,4429	5,6063	4,7295	1,3897	1,1787	1,0223	0,8789
300	7,5982	6,4591	5,5945	4,6932	1,3429	1,1299	0,9734	0,8299
400	7,6798	6,5179	5,6223	4,7233	1,3131	1,1064	0,9569	0,8152
500	7,7796	6,5900	5,6525	4,7297	1,3111	1,1030	0,9488	0,8022
600	7,8144	6,5931	5,6794	4,7496	1,2995	1,0923	0,9437	0,7960

Test SupF

L. obs.	99%	97,5%	95%	90%	10%	5%	2,5%	1%
25	264,669	109,521	59,051	34,391	7,108	6,262	5,628	4,986
50	22,932	18,356	15,700	13,491	5,878	5,271	4,787	4,272
75	16,514	14,693	13,312	11,945	5,426	4,850	4,374	3,878
100	15,818	14,244	13,016	11,727	5,243	4,645	4,173	3,704
125	15,866	14,278	13,027	11,707	5,140	4,542	4,100	3,631
150	15,914	14,423	13,196	11,868	5,153	4,555	4,095	3,607
200	16,320	14,697	13,370	11,953	5,080	4,475	3,998	3,534
300	16,698	15,059	13,663	12,155	5,057	4,454	3,983	3,498
400	17,226	15,360	13,887	12,360	5,063	4,445	3,973	3,504
500	17,623	15,736	14,155	12,506	5,109	4,487	4,014	3,520
600	17,665	15,759	14,155	12,575	5 117	4,493	4,027	3,530

Źródło: Obliczenia własne.

Literatura

ANDREWS D.W.K. (1990), Tests for Parameter Instability and Structural Change with Unknown Change Point, *Discussion Paper* 943, Yale University, Cowles Foundation for Research in Economics.

CHAREMZA W.W. i D.F. DEADMAN (1997), *New Directions in Econometric Practice, General to Specific Modelling, Cointegration and Vector Autoregression*, wyd. drugie, Edward Elgar, Cheltenham.

DAVID F.N. i N.L. JOHNSON (1954), Statistical treatment of censored data, part 1, fundamental formulae, *Biometrika* 41, str. 228–240.

DICKEY D.A. i W.A. FULLER (1979), Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association*, 84, str. 427–431.

ENGLE R.F. i C. W. J. GRANGER (1987), Co-integration and error correction: Representation, estimation and testing, *Econometrica* 55, str. 251––276.

HALDRUP N. (1994), Semiparametric tests for double unit roots, *Journal of Business & Economic Statistics* 12(1), str. 109–122.

HANSEN B.E. (1991), *Tests for Parameter Instability in Regressions with I(1) Processes*, Working Paper, University of Rochester.

HANSEN B.E. (1992a), Efficient estimation and testing of cointegrating vectors in the presence of deterministic trends, *Journal of Econometrics*, 53, str. 87–121.

HANSEN B.E. (1992), Tests for Parameter Instability in Regressions with I(1) Processes, *Journal of Business & Economic Statistics*, 10(3), str. 321–335.

MACKINNON (1991), *Critical Values for Cointegration Tests*, rozdział 13 pracy: *Long-Run Economic Relationships, Readings in Cointegration*, red. R.F. Engle i C.W.J. Granger, Oxford University Press, Oxford 1991, str. 267–276.

MACKINNON J. (1994), Approximate asymptotic distribution functions for Unit-root and Cointegration tests, *Journal of Business & Economic Statistics*, 12, str. 167–176.

MACKINNON J. (1996), Numerical distribution functions for unit root and cointegration tests, *Journal of Applied Econometrics*, 11(6), str. 601–618.

OULIARIS S. i P.C.B. PHILLIPS (1994), *COINT: GAUSS Procedures for Cointegrated Regressions*.

PAWŁOWSKI Z. (1980), *Statystyka matematyczna*, wyd. drugie, PWN, Warszawa.

PHILLIPS P.C.B. i B.E. HANSEN (1990), Statistical Inference in Instrumental Variables Regression with I(1) Processes, *Review of Economic Studies*, 57, 99–125.

SYCZEWSKA E.M. (1999), *Analiza relacji długookresowych: estymacja i weryfikacja*, nr 462 serii Monografie i Opracowania, Szkoła Główna Handlowa, Warszawa.

Hansen Tests as Cointegration Tests for I(2), I(1) Series

Abstract

Hansen tests, used for testing stability of cointegration relation, can be applied to cointegration regression in which dependent variable is I(1) and both explanatory variables are I(2) cointegrated to I(1). This kind of model describes e.g. real exchange rates. As a result of Monte Carlo experiment properties of test statistics are compared and tables of critical values are presented.

Jadwiga Orylska, Liwiusz Siemianowski
Zakład Informatyki Akademii Rolniczej w Szczecinie

Z badań nad modelowaniem i prognozowaniem wskaźników zanieczyszczenia Odry

Wstęp

Potrzeba racjonalnego korzystania z zasobów środowiska przyrodniczego, realizacja zrównoważonego rozwoju wymaga sprawnego, efektywnego systemu obserwacji środowiska z wykorzystaniem technologii informatycznych i metod matematycznych. Wydaje się że w funkcjonującym w Polsce Państwowym Monitoringu Środowiska wskazane jest szersze wykorzystanie metod matematycznych do sporządzania ocen i prognoz stanu środowiska (Orylska 1995; Orylska i Siemianowski 1994; Siemianowski 1997).

Celem referatu jest próba przedstawienia rezultatów modelowania i prognozowania wskaźników zanieczyszczenia Odry metodami matematycznymi.

Metodyka i materiał empiryczny

W referacie przedstawiono zastosowanie metod matematycznych do modelowania trendu i prognozowania jakości wody Odry z analizą sezonowości oraz współzależności.

Dane empiryczne dla celów badawczych udostępnił Wojewódzki Inspektorat Ochrony Środowiska w Szczecinie.

Badaniu poddano wybrane, istotne parametry, jak: temperatura wody, odczyn pH, biologiczne zapotrzebowanie tlenu (pięciodniowe) – BZT_5, chemiczne zapotrzebowanie tlenu – ChZT, stężenie tlenu rozpuszczonego, fosforanów, azotu ogólnego, chlorofil „a" oraz przepływ wody. Dla celów analizy obliczono średnie miesięczne. Modelowanie dynamiki i wahań przeprowadzono z zastosowaniem modeli trendu i współzależności z wahaniami sezonowymi.

Charakterystyka zanieczyszczeń wód płynących

Składnikami wód płynących obok substancji naturalnych skorupy ziemskiej (np. jony wapniowe, sodowe, magnezowe, siarczanowe, chlorkowe i inne) są różnego rodzaju zanieczyszczenia pochodzenia przyrodniczego oraz antropogennego. Zanieczyszczenie, a zarazem jakość wód opisywane są wieloma wskaźnikami fizykochemicznymi i biologicznymi.

Temperatura wywiera duży wpływ na biocenozy wód i zachodzące w nich procesy chemiczne. Poszczególne organizmy wodne cechuje określony zakres temperatur dla życia i rozwoju. Przekroczenie limitów może być szkodliwe. Nagła zmiana temperatury może również być śmiertelna, mimo nieprzekroczenia granicy letalnej. Wysoka temperatura przyspiesza przebieg procesów chemicznych i biochemicznych w wodzie, np. procesu nitryfikacji. Wzrost temperatury zmniejsza rozpuszczalność tlenu w wodzie, co może doprowadzić do deficytu tlenowego i spadku zdolności asymilacyjnej odbiornika (rzeki), tj. zmniejszenia ładunku ścieków możliwego do przyjęcia, oraz zwiększenia wrażliwości organizmów wodnych na działanie toksykantów.

Odczyn pH określa stopień zakwaszenia lub zasadowości. Odczyn wód naturalnych zależy od rodzaju podłoża i gleb w zlewni oraz od zanieczyszczeń (ścieki, opady atmosferyczne, spływy powierzchniowe). Wody powierzchniowe mają pH zwykle w przedziale 6,5–8,5. Znaczny wpływ na zakwaszanie wód powierzchniowych mają tzw. kwaśne deszcze będące wynikiem spalania węgla i ropy naftowej. Wiele gatunków ryb ginie już przy pH = 5,5. Zarówno wody zbyt kwaśne, jak i zbyt alkaliczne są szkodliwe. Szczególnie wody kwaśne mają niekorzystne działanie, powodują przyspieszoną korozję metali oraz wymywanie metali ciężkich i radionuklidów z osadów dennych.

Tlen rozpuszczony w wodzie ma podstawowe znaczenie dla wszelkich procesów zachodzących w wodach naturalnych, jest niezbędny do życia ryb i innych organizmów, jest zużywany przy rozkładzie substancji organicznych umożliwiając samooczyszczanie wód. Rozpuszczalność tlenu w wodzie maleje ze wzrostem temperatury i zasolenia. Niedotlenienie może wystąpić przy zbyt szybkim zużywaniu tlenu (np. z powodu nadmiernego zanieczyszczenie ściekami). Deficyt tlenowy (szczególnie duży to 30-procentowe nasycenie tlenem lub 2–3 mg tlenu na dm^3) powoduje śnięcie ryb, mogą powstać warunki beztlenowe (anaerobowe), objawiające się nieprzyjemnym odorem (powstaje siarkowodór).

Związki azotu dostają się wraz ze ściekami miejskimi i przemysłowymi, spływami powierzchniowymi i opadami atmosferycznymi. Duży udział zanieczyszczenia wód powierzchniowych azotem pochodzi od nie zawsze właściwego stosowania związków azotowych w rolnictwie (zbyt duże dawki nawozów).

Fosfor w wodach powierzchniowych występuje z przyczyn naturalnych: w wyniku wietrzenia i rozpuszczania minerałów fosforanowych, erozji gleby oraz w wyniku dopływu ścieków komunalnych i przemysłowych, spływów powierzchniowych i opadów atmosferycznych. W ściekach bytowych znajdują się znaczne ilości fosforanów ze środków piorących zawierających te związki. Biogenne związki azotu i fosforu intensyfikują rozwój fitoplanktonu (glonów).

Zanieczyszczenie wody związkami organicznymi jest określane za pomocą takich wskaźników, jak: BZT_5 – pięciodniowe biochemiczne zapotrzebowanie tlenu, utlenialność – ChZT(Mn), chemiczne zapotrzebowanie tlenu metodą manganową, oraz chemiczne zapotrzebowanie tlenu metodą chromianową – ChZT(Cr). Wymienione cechy obrazują zużywanie tlenu. Związki organiczne dostają się do wód ze ściekami przemysłowymi i komunalnymi (węglowodory, fenole). Pochodzą również ze spływów powierzchniowych z pól (chemikalia rolnicze – pestycydy), z opadów atmosferycznych oraz ze źródeł naturalnych. Duże stężenia związków organicznych są szkodliwe dla organizmów, pogarszają smak i zapach wody.

Chlorofil „a" – jeden z zielonych barwników roślin (odpowiedzialny za asymilację dwutlenku węgla w fotosyntezie), jest wskaźnikiem zawartości biomasy (fitoplanktonu) w wodzie i intensywności fotosyntezy.

Stężenie związków w wodzie związane jest również z przepływem wody. Wszystkie wskaźniki wykazują wahania sezonowe.

W Polsce wody rzeczne klasyfikowane są według czterostopniowej skali. Wody klasy I to wody nie zanieczyszczone, nadające się jako źródła wody pitnej. Wody klasy II nadają się do hodowli ryb i rekreacji, wody klasy III mogą być wykorzystane do celów przemysłowych i nawodnień rolniczych. Wody pozaklasowe (nie odpowiadające normatywom – n.o.n.) oznaczają wody nie nadające się do wykorzystania użytkowego. Normy wskaźników przedstawia tabela 1.

Tabela 1
Normy wybranych wskaźników klasyfikujące wodę do danej klasy

Cecha	Jednostka	Klasa I		Klasa II		Klasa III		n.o.n.
		od	do	od	do	od	do	
Tempratura wody	°C	pon. 22	22	22	26	pon. 26	26	pow. 26
Odczyn	pH	6,5	8,5	8,5	9	6	6,5	pon. 6 i pow. 9
Tlen rozpuszczony	mg O_2/dm^3	pow. 6	6	6	5	5	4	pon. 4
BZT_5	mg O_2/dm^3	0	4	4,0	8	8	12	pow. 12
ChZT-Mn	mg O_2/dm^3	0	10	10	20	20	30	pow. 30
Azot ogólny	mg N/dm^3	0	5	5	10	10	15	pow. 15
Fosforany	mg PO_4/dm^3	0	0,2	0,2	0,6	0,6	1	pow. 1
Chlorofil „a"	g/dm^3	0	10	10	20	20	30	pow. 30

Źródło: Opracowanie na podstawie Rozp. MOŚZNiL z 5 XI 1991, Dz. U. 116, poz. 503 z 1991 r.

Prognozowanie i modelowanie wskaźników jakości wody

Przedstawiono rezultaty zastosowania metod ekonometrycznych do modelowania dynamiki i wahań oraz prognozowania wskaźników zanieczyszczenia wód Odry (Zawadzki 1995; Raport o stanie... 1995).

Analizowano wybrane ważne wskaźniki jakości wody Odry, między innymi: odczyn pH, chlorofil „a", substancje biogenne.

Parametry równań modeli szeregu czasowego opisujących badane zmienne szacowano na podstawie 48 obserwacji średnich miesięcznych z okresu listopad 1990 – październik 1994.

Prognozy zostały wyznaczone dla okresu listopad 1994 – październik 1995. Oceny współczynników determinacji R^2 wskazywały na dobre dopasowanie wartości teoretycznych do empirycznych (przykładowo dla chlorofilu „a" $R^2 = 0,92$; dla fosforanów – 0,73, odczynu pH – 0,77).

Zastosowano model trendu ze stałą sezonowością o postaci:

$$Y_t = \alpha_1 t + \alpha_0 + \sum_{k=1}^{m} d_k Q_{kt} + U_t \qquad (I)$$

gdzie:
Y_t – zmienna objaśniana,

Q_{kt} – macierz zerojedynkowa przyjmująca wartości 1 w k-tym sezonie i wartość zero w pozostałych sezonach,
α_0, α_1 – parametry trendu,
d_k – parametry opisujące wahania sezonowe, k – indeks sezonów,
t – zmienna czasowa, U_t – składnik losowy.

Wykresy przedstawiają przebieg wartości wyznaczonych prognoz na tle wartości empirycznych (rys. 1–11). Współczynniki determinacji i błędy względne prognoz zestawiono w tabeli 2.

Tabela 2
Względne i przeciętne błędy prognoz badanych cech jakości wody Odry oraz wartości współczynnika determinacji R^2

t	Miesiąc, rok	Przepływ	Temperatura wody	Odczyn pH	Tlen rozpuszczony	BZT_5	ChZT (Mn)	Azot ogólny	Fosforany	Chlorofil „a"
49	Lis-94	0,11	0,39	0,003	0,05	0,03	0,03	0,17	0,25	0,86
50	Gru-94	0,17	0,66	0,02	0,06	0,32	0,01	0,18	0,02	0,42
51	Sty-95	0,28	0,46	0,02	0,10	0,25	0,09	0,06	0,08	0,67
52	Lut-95	0,16	1,15	0,01	0,01	0,20	0,002	0,21	0,18	0,84
53	Mar-95	0,12	0,17	0,03	0,05	0,11	0,03	0,09	0,44	1,30
54	Kwi-95	0,32	0,19	0,03	0,03	0,17	0,02	0,18	0,24	0,80
55	Maj-95	0,08	0,09	0,001	0,02	0,004	0,02	0,01	0,15	0,27
56	Cze-95	0,37	0,09	0,09	0,29	0,55	0,27	0,00	0,88	0,44
57	Lip-95	0,20	0,08	0,04	0,10	0,20	0,15	0,03	0,28	0,04
58	Sie-95	0,23	0,07	0,03	0,005	0,16	0,03	0,14	0,11	0,59
59	Wrz-95	0,18	0,09	0,04	0,15	0,27	0,13	0,21	0,13	0,11
60	Paź-95	0,24	0,23	0,03	0,19	0,39	0,13	0,05	0,21	0,26
Przeciętny błąd względny		**0,20**	**0,31**	**0,03**	**0,09**	**0,22**	**0,08**	**0,26**	**0,25**	**0,19**
R^2		**0,57**	**0,93**	**0,77**	**0,24**	**0,64**	**0,74**	**0,75**	**0,73**	**0,92**

Źródło: Opracowanie własne.

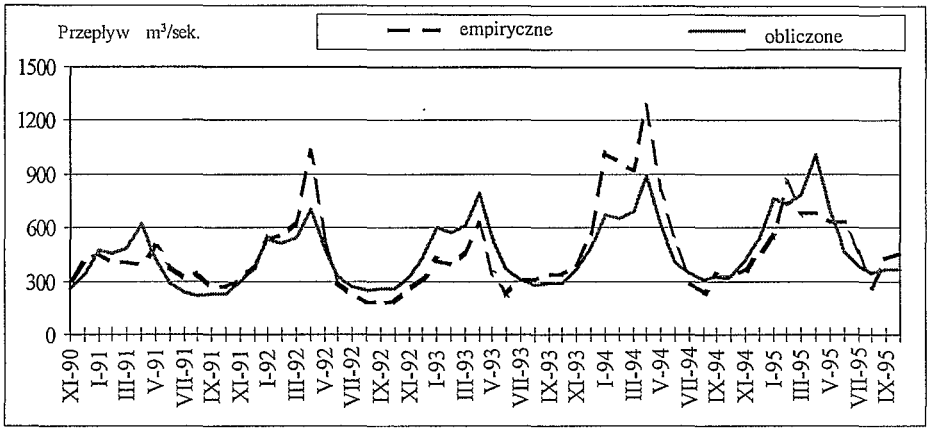
Parametry fizykochemiczne wody podlegają również wzajemnym współzależnościom (tab. 3). Przykładowo, wzrost temperatury, duży przepływ wody, wzrost zawartości fosforu i zakwit glonów (chlorofil) powodują wzrost zużywania tlenu obrazowanych przez BZT_5 i ChZT.

Silną dodatnią korelację (tab. 3) wykazują między innymi: temperatura wody z BZT_5 i ChZT, odczyn pH z BZT_5 i ChZT, chlorofil „a" z BZT_5 i ChZT.

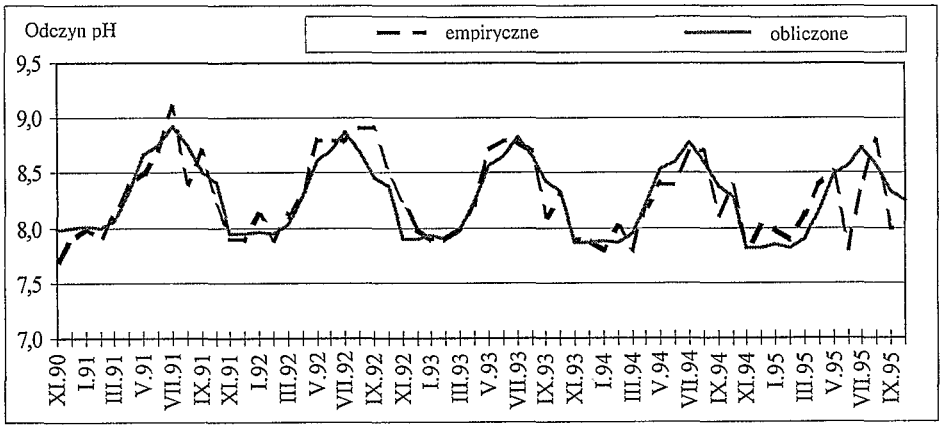
Tabela 3
Współczynniki korelacji badanych wskaźników jakości wody dla średnich miesięcznych z okresu listopad 1990 – październik 1995 w punkcie pomiarowo-kontrolnym na Odrze w Krajniku Dolnym

	Przepływ	Temperatura wody	Odczyn pH	Tlen rozpuszczony	BZT_5	ChZT (Mn)	Azot ogólny	Fosforany	Chlorofil „a"
Przepływ	1,000	–0,331	–0,385	0,265	–0,560	–0,349	0,634	–0,481	–0,521
Temperatura wody	–0,331	1,000	0,793	–0,451	0,675	0,819	–0,740	–0,393	0,752
Odczyn pH	–0,385	0,793	1,000	–0,090	0,832	0,878	–0,550	–0,333	0,880
Tlen rozpuszczony	0,265	–0,451	–0,090	1,000	–0,152	–0,240	0,484	–0,307	–0,105
BZT_5	–0,560	0,675	0,832	–0,152	1,000	0,816	–0,501	–0,059	0,840
ChZT (Mn)	–0,349	0,819	0,878	–0,240	0,816	1,000	–0,541	–0,301	0,891
Azot ogólny	0,634	–0,740	–0,550	0,484	–0,501	–0,541	1,000	0,022	–0,561
Fosforany	–0,481	–0,393	–0,333	–0,307	–0,059	–0,301	0,022	1,000	–0,207
Chlorofil „a"	–0,521	0,752	0,880	–0,105	0,840	0,891	–0,561	–0,207	1,000

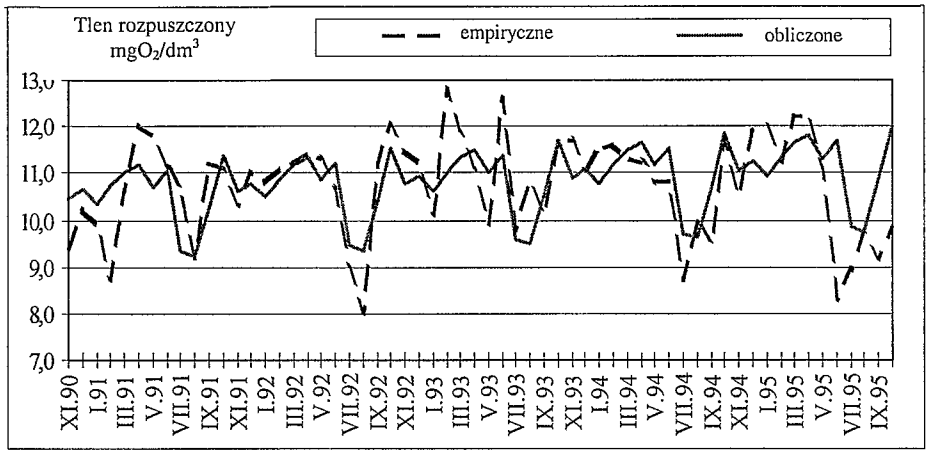
Źródło: Opracowanie własne.



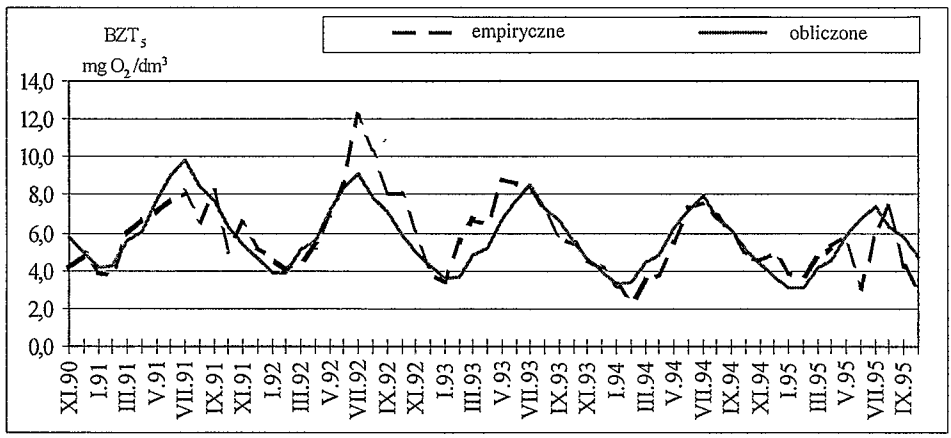
Rysunek 1
Wartości empiryczne i prognozowane przepływu wody Odry w Krajniku
Źródło: Opracowanie własne.



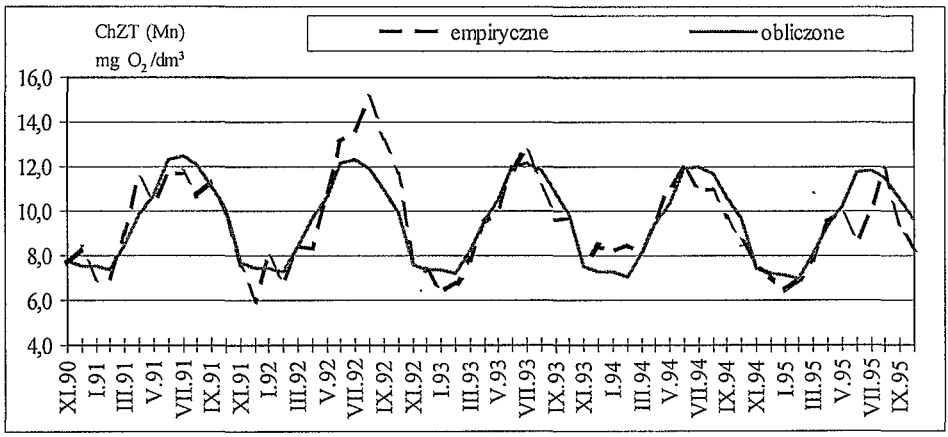
Rysunek 2
Wartości empiryczne i prognozowane odczynu pH wody Odry w Krajniku
Źródło: Opracowanie własne.



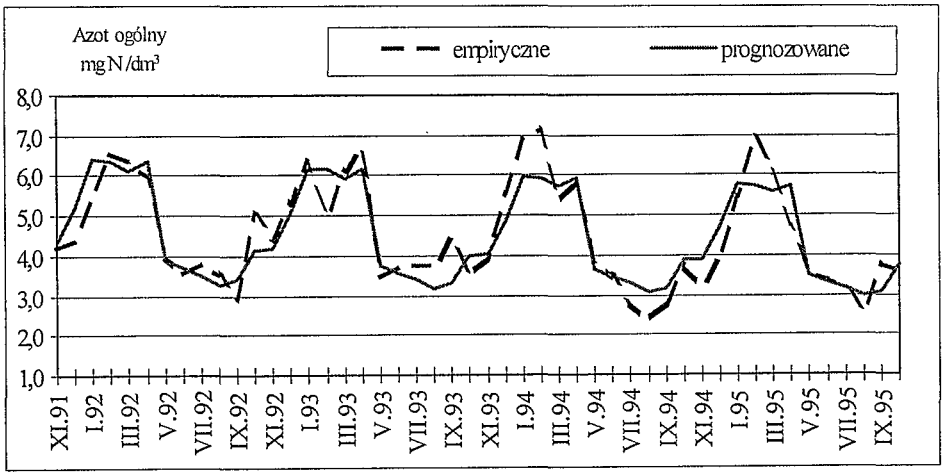
Rysunek 3
Wartości empiryczne i prognozowane tlenu rozpuszczonego wody Odry w Krajniku
Źródło: Opracowanie własne.



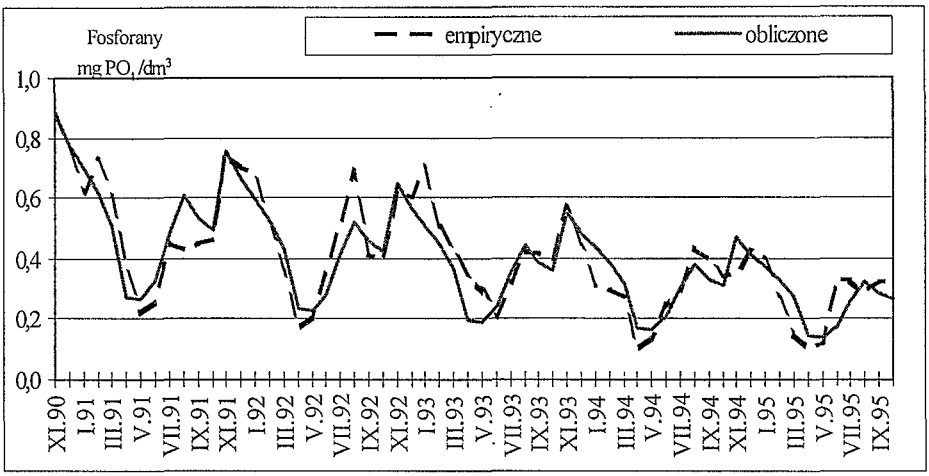
Rysunek 4
Wartości empiryczne i prognozowane BZT_5 wody Odry w Krajniku
Źródło: Opracowanie własne.



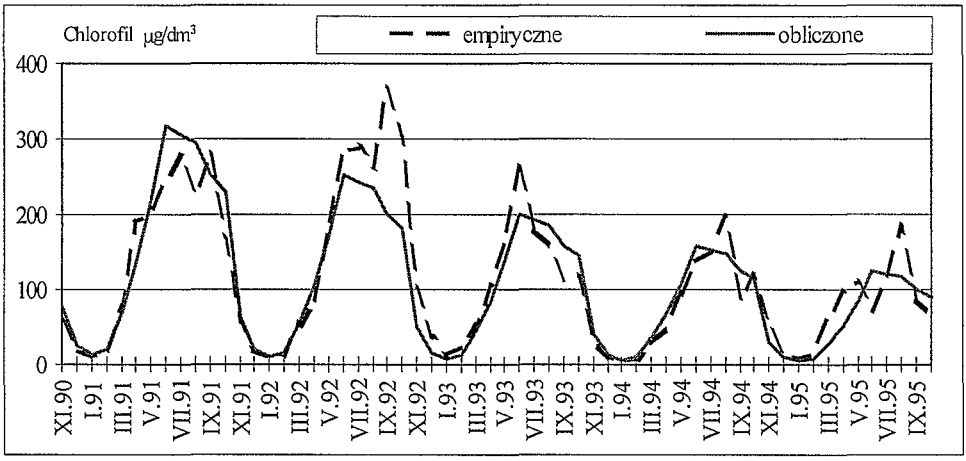
Rysunek 5
Wartości empiryczne i prognozowane ChZT(Mn) wody Odry w Krajniku
Źródło: Opracowanie własne.



Rysunek 6
Wartości empiryczne i prognozowane azotu ogólnego w wodzie Odry w Krajniku
Źródło: Opracowanie własne.



Rysunek 7
Wartości empiryczne i prognozowane fosforanów w wodzie Odry w Krajniku
Źródło: Opracowanie własne.



Rysunek 8
Wartości empiryczne i prognozowane chlorofilu „a" w wodzie Odry w Krajniku
Źródło: Opracowanie własne.

Celem uzyskania pełniejszego obrazu kształtowania się badanych cech jakości wody zastosowano ponadto poniższy model współzależności (II):

$$Y_t = \alpha_0 + \sum_{i=1}^{n} \alpha_i X_{it} + \sum_{k=1}^{m} d_{0k} Q_{kt} \quad \text{(II)},$$

w którym wprowadzono zmienne objaśniające – X_i.

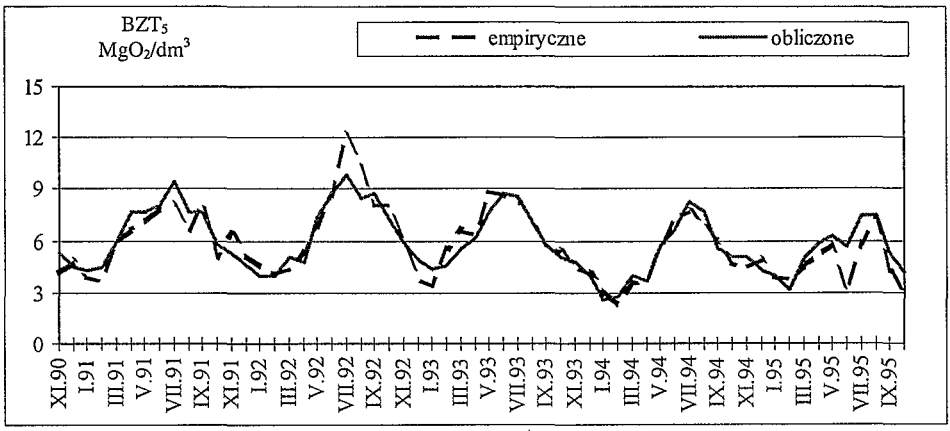
Wyznaczono prognozy zużywania tlenu. Zmienną endogeniczną modelu stanowiło biochemiczne zapotrzebowanie tlenu – BZT_5. Przeanalizowano trzy warianty modelu (II), wykorzystując jako zmienne egzogeniczne poniższe cechy:

a) przepływ i chlorofil „a",
b) przepływ i temperatura wody,
c) przepływ i odczyn pH.

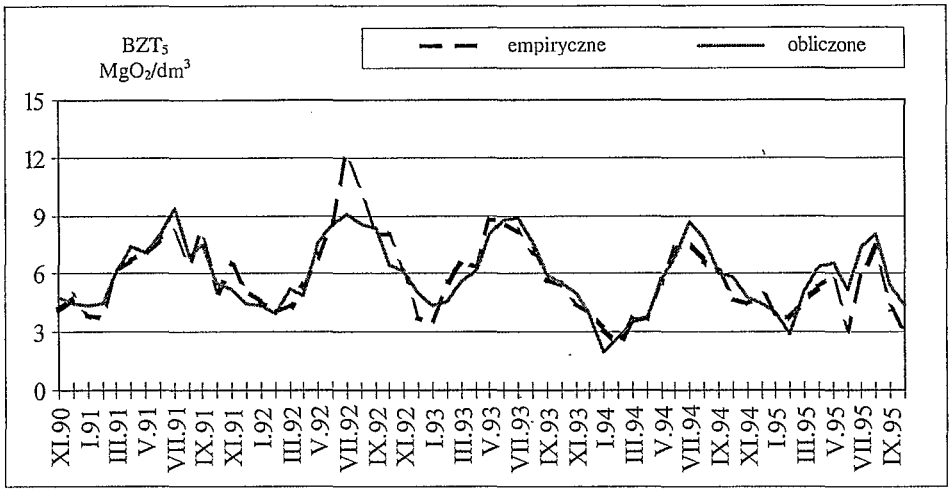
Otrzymane wyniki zilustrowano na wykresach (rys. 9–11).

Najlepsze wyjaśnienie zmienności BZT_5 uzyskano dla zmiennych objaśniających: przepływ i odczyn pH (model c, $R^2 = 0{,}79$). Pozostałe dwa warianty modelu również cechują się dobrym dopasowaniem (model a, $R^2 = 0{,}71$ i model b, $R^2 = 0{,}75$).

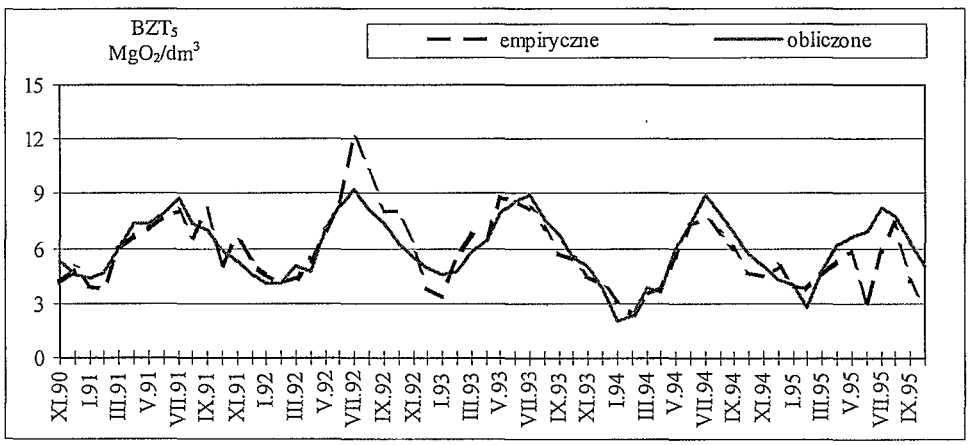
W tabeli 4 przedstawiono obliczone błędy względne prognoz i współczynniki determinacji modelu II.



Rysunek 9
Wartości BZT_5 w zależności od przepływu i chlorofilu „a" (model a)
Źródło: Opracowanie własne.



Rysunek 10
Wartości BZT_5 w zależności od przepływu i odczynu pH (model b)
Źródło: Opracowanie własne.



Rysunek 11
Wartości BZT_5 w zależności od przepływu i temperatury wody (model c)
Źródło: Opracowanie własne.

Tabela 4
Błędy względne prognoz i współczynniki determinacji modeli współzależności (II) cech jakości wody Odry w Krajniku Dolnym

Miesiąc, rok	Model a	Model b	Model c
Listopad 94	0,135	0,015	0,041
Grudzień 94	0,121	0,004	0,050
Styczeń 95	0,007	0,025	0,014
Luty 95	0,164	0,122	0,062
Marzec 95	0,089	0,044	0,058
Kwiecień 95	0,103	0,055	0,025
Maj 95	0,084	0,058	0,021
Czerwiec 95	0,886	0,218	0,249
Lipiec 95	0,256	0,111	0,104
Sierpień 95	0,019	0,034	0,037
Wrzesień 95	0,262	0,191	0,138
Październik 95	0,454	0,188	0,120
Średnie	0,215	0,089	0,077
R^2	0,71	0,75	0,79

Źródło: Opracowanie własne.

Wnioski

W niniejszym referacie wykazano przydatność metod ekonometrycznych w modelowaniu i prognozowaniu cech jakości wody. Przedstawiona procedura umożliwia przeanalizowanie kształtowania się kolejnych wskaźników w monitoringu Odry, wód Zalewu Szczecińskiego lub kolejnych rzek. Proponowane metody powinny znaleźć szersze zastosowanie w przygotowywaniu ocen i prognoz stanu środowiska w Państwowym Monitoringu Środowiska.

Literatura

1. ORYLSKA J. (red.), 1995: *Zastosowania informatyki w gospodarce – systemy informatyczne ochrony środowiska*, Akademia Rolnicza, Szczecin.
2. ORYLSKA J., SIEMIANOWSKI L., 1994: Wybrane systemy informatyczne w ochronie środowiska naturalnego. Materiały konferencyjne *Do-*

radztwo w ekorozwoju obszarów wiejskich, AR Szczecin – ART Olsztyn, Szczecin.
3. SIEMIANOWSKI L., 1997: Problemy obserwacji i prognozowania zanieczyszczenia terenów Odry z wykorzystaniem technologii informatycznych – rozprawa doktorska, Uniwersytet Szczeciński.
4. ZAWADZKI J., 1995: *Ekonometryczne metody prognozowania procesów gospodarczych*, Akademia Rolnicza, Szczecin.
5. *Raport o stanie środowiska w województwie szczecińskim*, Wojewódzki Inspektorat Ochrony Środowiska, Szczecin, 1995, praca zbiorowa p. red. T. Mutko.
6. Rozporządzenie Ministra Ochrony Środowiska, Zasobów Naturalnych i Leśnictwa z 5 XI 1991, Dz. U. 116, poz. 503 z 1991 r.

Application of selected methods for modelling and predicting the water quality parameters of Odra river

Abstract

This paper is an presenting of the trial of application the selected mathematical methods for modelling and predicting the water quality parametres of Odra river with analysis of seasonality and regression. The main following parameters ware modelled: water temperature, concentration of dissolved oxygen, biochemical oxygen demand – BOD5, chemical oxygen demand, concentrations of phosphates and nitrates, chlorophyll „a" and water flow. The monthly averages (of five hydrological years range) of mentioned parameters were calculated for analysing. Modelling and estimating the predictions of the water quality parameters of Odra river were performed with using the trend and regression model with seasonality analysis. The results point the asability of presented method at predicting the water quality parameters.

It seems that mathematical methods should be more often apllied in Polish State Environment Monitoring to prepare reports and assesments.

Barbara Kowalczyk
Instytut Ekonometrii SGH w Warszawie

Estymacja ilorazu dwóch średnich w kolejnym okresie badania

Wprowadzenie

Wiele spośród ważnych badań metodą reprezentacyjną, wykonywanych zarówno przez instytucje rządowe, jak i firmy komercyjne przeprowadzające analizę rynku, powtarzanych jest w regularnych odstępach czasu. Do badań takich wykonywanych przez GUS należą m.in.: „Badania aktywności ekonomicznej ludności", „Badania budżetów gospodarstw domowych" oraz „Spisy rolne". Częstym przedmiotem zainteresowania w tego typu badaniach jest estymacja ilorazu wartości globalnych (średnich) dwóch cech. Szacuje się m.in.: stosunek liczby pracujących mężczyzn do liczby pracujących kobiet, przeciętne wydatki i spożycie w gospodarstwie domowym na osobę, przeciętny plon pszenicy z ha, stosunek powierzchni przeznaczonej pod uprawę pszenicy do powierzchni pod uprawę żyta itd.

Przy powtarzaniu badania reprezentacyjnego powstaje pytanie: w jaki sposób wykorzystać informacje z okresu poprzedniego, aby zwiększyć efektywność oceny dla okresu badanego? Istnieje szeroko rozwinięta teoria dotycząca badań powtarzalnych, ale związana raczej z szacunkiem takich parametrów, jak średnia cechy (wartość globalna) na okres bieżący, różnica średnich w dwóch kolejnych okresach czy też średnia z kilku okresów badania. Problem estymacji ilorazu wartości globalnych (średnich) dwóch cech w przypadku badań powtarzalnych jest bardziej skomplikowany i rzadziej poruszany w literaturze, choć z praktycznego punktu widzenia na pewno nie mniej ważny.

Schemat losowania dwustopniowego

Zakładamy, że N-elementowa populacja została podzielona na M rozłącznych, niepustych podzbiorów zwanych jednostkami losowania pierwszego stopnia (lps) (w praktyce jednostkami lps są np. rejony statystyczne czy terenowe punkty badań). Jednostka losowania pierwszego stopnia o numerze h

zawiera N_h jednostek badania, nazywanych jednostkami losowania drugiego stopnia (lds) (dla przykładu w „Badaniach budżetów gospodarstw domowych" jednostkami lds są mieszkania). Losowanie dwustopniowe polega na tym, że najpierw losujemy spośród M jednostek lps m jednostek według wybranego schematu. Na drugim stopniu losowanie przeprowadza się tylko w wylosowanych wcześniej jednostkach lps, przy czym zakłada się, że liczby jednostek lds n_h losowanych z jednostek lps ustalane są przed losowaniem.

Wartość cechy Y u j-tej jednostki lds należącej do h-tej jednostki lps oznaczamy przez Y_{hj} (h = 1, 2, ..., M, j = 1, 2, ..., N_h). Przyjmujemy poza tym następujące oznaczenia:

- wartość globalna cechy Y w h-tej jednostce lps: $Y_h = \sum_{j=1}^{N_h} Y_{hj}$,
- wariancja cechy Y dla h-tej jednostki lps:

$$S_{2h}^2 = \frac{1}{N_h - 1} \sum_{j=1}^{N_h} (Y_{hj} - \frac{1}{N_h} Y_h)^2,$$

- wartość globalna cechy Y dla całej populacji: $Y = \sum_{h=1}^{M} Y_h$,
- wartość cechy Y u jednostki lds wylosowanej w i-tym ciągnieniu z jednostki lps wylosowanej w g-tym ciągnieniu: y_{gi}.

W artykule rozpatrujemy schemat losowania dwustopniowego: 1-lppxzz, 2-lpbz. Skrót 1-lppxzz oznacza, że na pierwszym stopniu stosowany jest schemat: losowanie z prawdopodobieństwami proporcjonalnymi do wartości cechy X ze zwracaniem (lppxzz), tzn. prawdopodobieństwo wyciągnięcia h-tej jednostki w pojedynczym ciągnieniu wynosi $p_h = \frac{X_h}{\sum_{h=1}^{M} X_h}$, natomiast na drugim stopniu stosowany jest schemat: losowanie proste bez zwracania (lpbz). Cechą dodatkową w praktyce jest często wielkość jednostki lps mierzona liczbą jednostek lds w niej zawartych.

Schemat losowania dwustopniowego należy do najczęściej stosowanych schematów w praktyce badań reprezentacyjnych. Poza tym, losowanie jednostopniowe jest szczególnym przypadkiem wyżej wymienionego, tak więc przedstawiona w artykule teoria przenosi się również na ten schemat losowania.

Jeżeli próba została wylosowana zgodnie ze schematem losowania dwustopniowego: 1-lppxzz, 2-lpbz, to estymatorem nieobciążonym wartości globalnej Y jest statystyka

$$y=\frac{1}{m}\sum_{g=1}^{m}\frac{N_g}{p_g n_g}\sum_{i=1}^{n_g}y_{gi}=\frac{1}{m}\sum_{g=1}^{m}\frac{N_g}{p_g}\bar{y}_g,$$

a jej wariancja dana jest wzorem

$$D^2(y)=\frac{1}{m}\left[\sum_{h=1}^{M}p_h\left(\frac{Y_h}{p_h}-Y\right)^2+\sum_{h=1}^{M}N_h(N_h-n_h)\frac{S_{2h}^2}{p_h n_h}\right].$$

Natomiast kowariancja

$$Cov(y,z)=\frac{1}{m}\left[\sum_{h=1}^{M}p_h\left(\frac{Y_h}{p_h}-Y\right)\left(\frac{Z_h}{p_h}-Z\right)+\sum_{h=1}^{M}N_h(N_h-n_h)\frac{S_{2hyz}}{p_h n_h}\right],$$

gdzie

$$S_{2hyz}=\frac{1}{N_h-1}\sum_{j=1}^{N_h}(Y_{hj}-\frac{1}{N_h}Y_h)(Z_{hj}-\frac{1}{N_h}Z_h).$$

W dalszym ciągu pracy stosujemy oznaczenia:

$$S^2(Y)=\left[\sum_{h=1}^{M}p_h\left(\frac{Y_h}{p_h}-Y\right)^2+\sum_{h=1}^{M}N_h(N_h-n_h)\frac{S_{2h}^2}{p_h n_h}\right],$$

$$C(Y,Z)=\left[\sum_{h=1}^{M}p_h\left(\frac{Y_h}{p_h}-Y\right)\left(\frac{Z_h}{p_h}-Z\right)+\sum_{h=1}^{M}N_h(N_h-n_h)\frac{S_{2hyz}}{p_h n_h}\right].$$

$$\rho(Y,Z)=\frac{C(Y,Z)}{S(Y)S(Z)}=\frac{Cov(y,z)}{D(y)D(z)}.$$

Rozważamy obecnie problem, w którym badanie reprezentacyjne dotyczące tej samej populacji przeprowadzane jest dwukrotnie. W pierwszym okresie losujemy m jednostek lps (następnie z każdej jednostki odpowiednią liczbę jednostek lds). W drugim okresie pewną część próby pozostawiamy do badania, a dokładniej pozostawiamy pm ($0 \leq p \leq 1$, pm $\in$ N) jednostek lps (wraz z wylosowanymi jednostkami lds), natomiast qm = (1 – p)m jednostek lps dolosowujemy z wyjściowej populacji, tj. spośród M jednostek lps i dalej z nich losujemy jednostki lds.

Interesuje nas odpowiedź na pytanie, jaką część próby (tj. frakcję jednostek lps) należy pozostawić do badania w drugim okresie oraz jak wykorzystać informacje z okresu poprzedniego w celu zwiększenia efektywności estymacji na okres bieżący.

Przyjmujemy następujące oznaczenia:

– estymator wartości globalnej Y_t w okresie t (t = 1, 2) obliczony na podstawie wspólnej dla obu okresów części próby:

$$y_{tp} = \frac{1}{mp} \sum_{g=1}^{mp} \frac{N_g}{p_g} \bar{y}_{tg},$$

– estymator wartości globalnej Y_t w okresie t (t = 1, 2) obliczony na podstawie części próby badanej tylko w jednym okresie:

$$y_{tq} = \frac{1}{mq} \sum_{g=mp+1}^{m} \frac{N_g}{p_g} \bar{y}_{tg}.$$

Proponowane estymatory

Klasyczny estymator $R_2 = \frac{Y_2}{X_2} = \frac{\bar{Y}_2}{\bar{X}_2}$ ilorazu wartości globalnych (średnich) dwóch cech w drugim okresie badania:

$$t = \frac{y_2}{x_2}.$$

Estymator ten nie wykorzystuje informacji z okresu poprzedniego. Przybliżony błąd średniokwadratowy tego estymatora wynosi:

$$MSE(t) \approx \frac{R_2^2}{m}\left[\frac{S^2(Y_2)}{Y_2^2}+\frac{S^2(X_2)}{X_2^2}-2\frac{C(Y_2,X_2)}{Y_2X_2}\right]=\frac{R_2^2G_2}{m},$$

gdzie

$$G_i = V^2(Y_i)+V^2(X_i)-2V(Y_i)V(X_i)\rho(Y_i,X_i),$$

$$V(Y_i)=\frac{S(Y_i)}{Y_i}, \quad V(X_i)=\frac{S(X_i)}{X_i}, \quad R_i=\frac{Y_i}{X_i}, \qquad i = 1, 2.$$

Estymator R_2 zaproponowany przez Tripathiego i Sinhę [2]:

$$t_{TS} = Qr'_{2p} + (1-Q)r_{2q},$$

gdzie

$$r_{2q}=\frac{y_{2q}}{x_{2q}}, \qquad r'_{2p}=\frac{y_{2p}+b(y_1-y_{1p})}{x_{2p}+b^*(x_1-x_{1p})},$$

$$b=\frac{C(Y_2,Y_1)}{S^2(Y_1)}, \qquad b^*=\frac{C(X_2,X_1)}{S^2(X_1)}.$$

Q jest tak dobrane, żeby minimalizowało błąd średniokwadratowy MSE(t_{TS}), tzn.

$$Q=\frac{MSE(r_{2q})}{MSE(r'_{2p})+MSE(r_{2q})}.$$

Przybliżony błąd średniokwadratowy tego estymatora wynosi:

$$MSE(t_{TS}) \approx \frac{R_2^2G_2}{m}\frac{G_2-qF}{G_2-q^2F}=\frac{R_2^2G_2}{m}\frac{1-q\frac{F}{G_2}}{1-q^2\frac{F}{G_2}},$$

gdzie

$$F = V^2(Y_2)\rho^2(Y_2,Y_1)+V^2(X_2)\rho^2(X_2,X_1)-2V(Y_2)V(X_2)\rho(X_2,X_1)\rho(Y_2,X_1)$$
$$-2V(Y_2)V(X_2)\rho(Y_2,Y_1)\rho(X_2,Y_1)+2V(Y_2)V(X_2)\rho(Y_2,Y_1)\rho(X_2,X_1)\rho(Y_1,X_1)$$

Optymalna wartość q (minimalizująca MSE(t_{TS})) przedstawia się następująco:

$$q_{opt} = 1, \quad gdy \quad F < 0,$$

$$q_{opt} = \frac{1}{1+\sqrt{1-\frac{F}{G_2}}}, \quad gdy \quad F \geq 0.$$

Esrtymator R_2 zaproponowany przez Okafora i Arnaba [1]:

$$t_{OA} = \frac{y_2'}{x_2'} = \frac{Q_y y_{2p}' + (1-Q_y) y_{2q}}{Q_x x_{2p}' + (1-Q_x) x_{2q}},$$

gdzie

$$y_{2p}' = y_{2p} + b_y(y_1 - y_{1p}), \quad x_{2p}' = x_{2p} + b_x(x_1 - x_{1p}).$$

Autorzy zaproponowali przyjąć takie wartości współczynników Q_y oraz b_y, które minimalizują $D^2(y_2)$ i analogicznie takie Q_x oraz b_x, które minimalizują $D^2(x_2)$, tzn.

$$b_z = \frac{C(Z_2, Z_1)}{S^2(Z_1)},$$

$$Q_z = \frac{D^2(z_{2q})}{D^2(z_{2q}) + D^2(z_{2p}')}, \qquad z = y,\ x.$$

Autorzy założyli, że

(i) $$\rho(y_1, y_2) = \rho(x_1, x_2) = \rho_0$$

i przy tym założeniu przyjęli $q = \frac{1}{1+\sqrt{1-\rho_0^2}}$, które minimalizuje jednocześnie $D^2(y_2')$ oraz $D^2(x_2')$.

Przyjmując dalsze założenia

(ii) $\rho(y_i, x_j) = \rho, \quad i, j = 1, 2,$

(iii) $V(Y_2) = V(X_2) = V_2$

autorzy przedstawili wzór na przybliżony błąd średniokwadratowy swojego estymatora.

$$MSE_{opt}(t_{OA}) \approx \frac{R_2^2 V_2^{\,2}}{m} \frac{1}{\sqrt{1-\rho_0^2}} \left\{ (1+\sqrt{1-\rho_0^2})(1-\rho) + \rho_0(\rho-\rho_0) \right\}.$$

Zaproponowana nowa postać estymatora R_2:

$$t_A = Q r_{2p}^{"} + (1-Q) r_{2q},$$

gdzie

$$r_{2q} = \frac{y_{2q}}{x_{2q}}, \qquad r_{2p}^{"} = r_{2p} + \tilde{b}(r_1 - r_{1p}) = \frac{y_{2p}}{x_{2p}} + \tilde{b}\left(\frac{y_1}{x_1} - \frac{y_{1p}}{x_{1p}}\right),$$

$$\tilde{b} = \frac{E(r_2 - R_2)(r_1 - R_1)}{E(r_1 - R_1)^2}.$$

Wykonując odpowiednie obliczenia, otrzymujemy:

$$\tilde{b} = \frac{I_Y}{I_X} \frac{H}{G_1},$$

gdzie

$$H = V(Y_1)V(Y_2)\rho(Y_1, Y_2) + V(X_1)V(X_2)\rho(X_1, X_2) + \\ -V(Y_1)V(X_2)\rho(Y_1, X_2) - V(Y_2)V(X_1)\rho(Y_2, X_1),$$

$$I_Y = \frac{Y_2}{Y_1}, \quad I_X = \frac{X_2}{X_1}.$$

Q wybieramy tak, aby minimalizowało $MSE(t_A)$, tzn.

$$Q = \frac{MSE(r_{2q})}{MSE(r_{2p}^{"}) + MSE(r_{2q})}.$$

Błąd średniokwadratowy zaproponowanego przeze mnie estymatora wynosi:

$$MSE(t_A) = \frac{R_2^2 G_2}{m} \frac{1 - q\frac{H^2}{G_1 G_2}}{1 - q^2 \frac{H^2}{G_1 G_2}}.$$

$MSE(t_A)$ osiąga najmniejszą wartość dla

$$q_{opt} = \frac{1}{1 + \sqrt{1 - \frac{H^2}{G_1 G_2}}}.$$

Zatem optymalna frakcja jednostek lps, które powinny być pozostawione do badania w okresie następnym, wynosi: $p_{opt} = 1 - q_{opt}$.

Podstawiając q_{opt} do wzoru na błąd średniokwadratowy otrzymujemy:

$$MSE_{opt}(t_A) \approx \frac{R_2^2 G_2}{m} \frac{1}{2}\left(1 + \sqrt{1 - \frac{H^2}{G_1 G_2}}\right).$$

Estymator zaproponowany przeze mnie jest trochę bardziej skomplikowany w konstrukcji niż estymatory proponowane wcześniej (trudno jest oszacować $\tilde{b}$). Ma on jednak tę zaletę, iż wszystkie współczynniki liczone były tak, aby minimalizowały błąd średniokwadratowy całego estymatora (w przypadku estymatora Tripathiego-Sinha współczynniki b, b^* nie spełniały tego kryterium).

Porównanie estymatorów oraz końcowe wnioski

W tej części opracowania założymy spełnienie warunków (i) – (iii). Dla różnych wartości ρ, ρ_0 porównamy efektywność estymatora zaproponowanego przeze mnie z efektywnością pozostałych estymatorów wykorzystujących informacje z okresu poprzedniego.

$MSE_{opt}(t_A)/MSE_{opt}(t_{TS})$:

ρ_0 \ ρ	0,1	0,3	0,5	0,7	0,9
0,1	1,000	0,979	0,800	--	--
0,3	0,998	1,000	0,958	--	--
0,5	0,999	0,988	1,000	0,873	--
0,7	0,9996	0,995	0,976	1,000	--
0,9	0,9999	0,999	0,995	0,980	1,000

$MSE_{opt}(t_A)/MSE_{opt}(t_{OA})$:

ρ_0 \ ρ	0,1	0,3	0,5	0,7	0,9
0,1	0,997	0,963	0,767	--	--
0,3	0,998	0,976	0,882	--	--
0,5	0,999	0,984	0,928	0,687	--
0,7	0,999	0,989	0,954	0,833	--
0,9	0,9995	0,994	0,974	0,910	0,607

Jak widać, estymator skonstruowany jako estymator liniowy od r_{2p}, r_{2q}, r_1, r_{1p} okazał się nieco lepszy od estymatorów proponowanych wcześniej (-- oznacza, że dana kombinacja ρ, ρ_0 jest niedopuszczalna). Zobaczymy teraz, jaki jest zysk na efektywności przy zastosowaniu estymatora t_A w porównaniu z zastosowaniem estymatora klasycznego, który nie wykorzystuje informacji z okresu poprzedniego.

$100\% [MSE(t) - MSE_{opt}(t_A)]/MSE_{opt}(t_A)$:

ρ_0 \ ρ	0,1	0,3	0,5	0,7	0,9
0,1	0,0	2,1	**25,0**	--	--
0,3	1,3	0,0	4,4	--	--
0,5	5,5	2,1	0,0	**14,6**	--
0,7	**14,6**	9,9	4,4	0,0	--
0,9	**37,2**	**32,0**	**25,0**	**14,6**	0,0

W zaznaczonych rubrykach (gdy odpowiednie współczynniki korelacji, jak można było się spodziewać, są wysokie) zysk na efektywności jest znaczny i przekracza 10%.

Na koniec powiemy jeszcze kilka słów o estymatorze zaproponowanym przez Okafora i Arnaba. Jego zaletą jest prosta konstrukcja. Jak widzieliśmy, okazał się on jednak mniej efektywny od estymatora zaproponowanego przeze mnie. W niektórych przypadkach okazuje się on mniej efektywny nawet od estymatora klasycznego t, który żadnych informacji z okresu poprzedniego nie wykorzystuje.

$MSE_{opt}(t_{OA})/MSE(t)$:

ρ_0 \ ρ	0,1	0,3	0,5	0,7	0,9
0,1	**1,003**	**1,017**	**1,043**	--	--
0,3	0,989	**1,024**	**1,087**	--	--
0,5	0,949	0,995	**1,077**	**1,270**	--
0,7	0,873	0,920	**1,004**	**1,200**	--
0,9	0,729	0,762	0,821	0,959	**1,647**

Literatura

1. OKAFOR F.C., ARNAB R., 1987: Some strategies of two-stage sampling for estimating population ratios over two occasions. *Austral. J. Statist.* 29, 128–142.
2. TRIPATHI T.P., SINHA S.K.P., 1976: Estimation of ratio on successive occasions. Paper presented at the „Symposium on Recent Developement in Survey Methodology" held at Indian Statistical Institute.

Estimation of population ratios on the most recent occasion

Abstract

The paper concerns survey sampling on repeated occasions. We discuss the estimation of the ratio of two population totals (means) on the second occasion using the two stage sampling. We present three estimators (including one new), based on informations from previous and recent occasions, and we compare the estimators with the ordinary ratio estimator.

Aleksander Busłowski
Katedra Metod Ilościowych i Informatyki WSE w Białymstoku

Współczynniki techniczne a stabilność rozwiązania optymalnego zadania programowania liniowego

Wprowadzenie

Brak prostych reguł (wzorów) rządzących zmianami wartości współczynników technicznych znajdujących się przy bazowych zmiennych decyzyjnych w warunkach napiętych przy decyzji optymalnej stanowi istotną przeszkodę w kompleksowej analizie zmian wartości parametrów wejściowych zadania programowania liniowego. Daje się to szczególnie zauważyć w podręcznikach do ekonometrii (badań operacyjnych), gdzie pomijana jest analiza zmian wartości współczynników technicznych [11, s. 278–293; 12; 13, s. 190–194]. Również programy komputerowe służące rozwiązywaniu zagadnień programowania liniowego nie oferują opcji analizy zmienności współczynników technicznych. Natomiast praktyka konstruowania modeli programowania liniowego wskazuje, że często faktyczne wartości współczynników funkcji celu i wyrazów wolnych są włączane do grupy współczynników przy zmiennych w warunkach ograniczających. Dodać również należy, iż współczynniki techniczne stanowią najliczniejszą grupę parametrów wejściowych każdego zadania programowania liniowego.

Dotychczas nie stawiano pytań, czy istnieje praktyczny sens badania zmian wartości wszystkich współczynników przy zmiennych w warunkach ograniczających[1]. Wiele przesłanek wskazuje na to, że istnieją parametry, których wartości – niezależnie od zmian w sytuacji decyzyjnej – są zawsze stałe (niezmienne).

Przedstawiane w literaturze pojęcie stabilności rozwiązania optymalnego [9, s. 169; 14, s. 225] wydaje się także zbyt ogólnikowe. W tej kwestii również zaproponowano pewne uściślenia i uszczegółowienia.

---

[1]Dotyczy to także pozostałych parametrów zadania programowania liniowego.

Celem niniejszego opracowania jest rekonstrukcja koncepcji analizy stabilności rozwiązania optymalnego w odniesieniu do reguł zmian wartości współczynników technicznych przy bazowych zmiennych decyzyjnych w warunkach napiętych przy decyzji optymalnej. W opracowaniu postawiono hipotezę, iż nieliniowe formuły zmienności tych współczynników można wyprowadzić z liniowych reguł zmian wartości współczynników funkcji celu i wyrazów wolnych. Sformułowanie tej hipotezy oparto na ścisłych związkach współczynników technicznych z wyrazami wolnymi i współczynnikami funkcji celu w każdym modelu programowania liniowego oraz na wynikających stąd analogicznych skutkach zmian wartości tych parametrów wejściowych dla uzyskanego rozwiązania optymalnego. Weryfikacja powyższej hipotezy wymagała dostosowania metodyki analizy dobrze opracowanych formuł zmienności współczynników funkcji celu i wyrazów wolnych w celu ich wykorzystania w analizie zmian wartości współczynników technicznych przy bazowych zmiennych decyzyjnych.

Pojęcie współczynnika przy zmiennej w warunku ograniczającym

W literaturze ekonomiczno-rolniczej spotyka się mnogość określeń i definicji pojęcia współczynnika [4, s. 382–383 i 853; 5, s. 24; 10, s. 112; 12, s. 24–25; 20, s. 236]. Wynika to z jego usytuowania w modelu, a w szczególności z charakteru warunku ograniczającego, w którym on występuje. Większość definicji współczynnika podkreśla jednostkowe zużycie (normy) środków ograniczających przez poszczególne działalności.

Jednakże nawet pobieżna analiza większości modeli wejściowych programowania liniowego wskazuje, że znajduje się tam wiele współczynników, które nie mają nic wspólnego ze zużyciem, z produkcją czy proporcjami. Przedstawiają one natomiast pewne zależności logiczne w modelu i przyjmują tylko trzy wartości: –1, 0 i +1. Nie należy wykluczać, że zużycie jakiegoś środka może przyjąć którąś z tych wartości, jednakże analiza modelu zawsze da nam odpowiedź, czy jest to zależność techniczna, czy logiczna. Przykładowo: jeżeli współczynnik +1 znajduje się przy zmiennej decyzyjnej „liczba krów" i w warunku ograniczającym „bilans pasz treściwych", to będzie współczynnikiem technicznym[2]. Ale jeśli w tym samym warunku ograniczającym, przy zmiennej decyzyjnej „ilość mieszanki KBW z zakupu" znajdzie się współczynnik –1, to

[2]Jego wartość jest potencjalnie zmienna w zależności od stosowanej technologii żywienia krów.

będzie on współczynnikiem logicznym, ponieważ niezależnie od zapotrzebowania na mieszankę KBW, ta wartość parametru (–1) będzie zawsze je równoważyć. Należy przy tym dodać, że występujące w modelu wejściowym – w sposób jawny lub też w postaci ukrytej – zmienne swobodne i sztuczne mają współczynniki wyłącznie logiczne.

Pewne wątpliwości mogą się pojawić w przypadku zera. Czy należy to odczytywać jako zerowy nakład (zużycie) – zależność techniczna, czy też jako brak nakładu (zużycia) – zależność logiczna? Mimo tożsamości zapisu w modelu obu określeń – należy opowiedzieć się za drugim, ze względu na częstość występowania.

Celem uniknięcia wymienionych wątpliwości należałoby w przypadku współczynników o charakterze zależności technicznej używać wartości nieznacznie różniących się od 0 i ±1. Nie zmieniłoby to późniejszych wyników, a dawałoby – bez wnikania w opis warunków ograniczających – jednoznaczną i wizualną charakterystykę zależności. I chyba najistotniejszy aspekt sprawy – współczynniki te byłyby rozróżnialne przez komputer.

Reasumując – współczynniki przy zmiennych w warunkach ograniczających należałoby podzielić na dwie grupy:

a) techniczne, tzn. zmienne w sensie ilościowym;
b) logiczne, tzn. niezmienne (stałe), które przyjmują tylko wartości 0 i ±1.

Podział ten wynika stąd, iż analiza stabilności powinna dotyczyć jedynie współczynników zaliczonych do grupy a), ponieważ znajdujące się w tych miejscach modelu parametry wejściowe – w zależności od sytuacji decyzyjnej – mogą przyjmować różne wartości, czego nie można powiedzieć o współczynnikach zaliczonych do grupy b). Należy również stwierdzić, że z rachunkowego punktu widzenia można badać także zmienność współczynników z grupy b), czyli pełną ich zbiorowość, jednakże nie miałoby to żadnego znaczenia praktycznego. Wynika to stąd, że ta grupa parametrów ze swej istoty jest niezmienna i przedstawia uwarunkowania wewnętrzne modelu, nie zaś zależności zmienne w sensie ilościowym.

Pojęcie stabilności

Przedmiotem analizy stabilności rozwiązania optymalnego zadania programowania liniowego jest badanie wpływu zmian wartości parametrów wejściowych zadania na uzyskane bazowe rozwiązanie optymalne. Analiza stabilności jest częścią tzw. analizy pooptymalizacyjnej (postoptymalizacyjnej) [1, s. 332; 14, s. 225; 16, s. 222; 17; 18].

Stabilność rozwiązania jest warunkiem dostatecznym jego optymalności. Oznacza to, iż zmiany wartości parametrów zadania nie mogą naruszyć stanu optymalności. Naruszenie stabilności rozwiązania jest jednoznaczne z utratą jego optymalności. Przywracanie naruszonej stabilności może być również przedmiotem analizy pooptymalizacyjnej, o ile do tego nie wykorzystuje się dodatkowego postępowania iteracyjnego algorytmem simpleksowym.

Z warunków stabilności wynika, że analizie tej nie podlega kształtowanie się wartości funkcji celu, ponieważ jest ona kategorią wyłącznie wynikową, a poziom jej wartości musi być zawsze optymalny (warunek konieczny). Natomiast warunki dodatkowe (wahania wartości zmiennych bazowych) dają możliwość ustalenia kilku rodzajów (poziomów) stabilności. Proponuje się następujące:

1) stabilność wartości zmiennych;
2) stabilność wartości zmiennych decyzyjnych;
3) stabilność bazy (struktury zmiennych bazowych, zmiennych niebazowych).

W pierwszym przypadku zmiany wartości parametrów wejściowych nie mogą mieć wpływu na wartości zmiennych bazowych. Dotyczy to wszystkich zmiennych, które mogą stanowić skład bazy rozwiązania optymalnego, a więc zmiennych decyzyjnych i swobodnych.

Drugi przypadek ogranicza niezmienność wartości wyłącznie do zmiennych decyzyjnych. Jak wiadomo, są one zasadniczymi elementami rozwiązania. Oznacza to, że zmienne, które są związane z warunkami luźnymi przy decyzji optymalnej, czyli bazowe zmienne swobodne, mogą zmieniać swoje wartości, a wystarcza jedynie fakt ich obecności w bazie rozwiązania.

Trzeci rodzaj stabilności zachowuje jedynie wartości zmiennych niebazowych, które w bazowym rozwiązaniu optymalnym są zawsze równe zeru. Natomiast zmienne bazowe mogą przyjmować dowolne wartości nieujemne.

Oceniając poszczególne przypadki stabilności rozwiązania, należy stwierdzić, że najbardziej pożądanym jest przypadek 1). Sytuacja 2) gwarantuje również dotrzymanie najistotniejszych wartości rozwiązania, tj. zmiennych decyzyjnych, więc z praktycznego punktu widzenia niewiele różni się od sytuacji 1). Przypadek 3) – ze względu na możliwości zmiany wartości wszystkich zmiennych bazowych – może mieć znaczną przydatność praktyczną w dostosowywaniu rozwiązania problemu decyzyjnego do wymogów rzeczywistości.

Z powyższego wynika, że rodzaje stabilności rozwiązania optymalnego mają układ hierarchiczny. Mimo tego, przekroczenie zakresu możliwej zmiany wartości parametru wejściowego nie da możliwości przejścia na niższy poziom stabilności, lecz zawsze pociągnie za sobą zmiany w strukturze bazy. Oznacza to,

iż w bazie rozwiązania pojawi się co najmniej jedna nowa zmienna[3]. Oznacza to również, że określony parametr wejściowy – zależnie od jego usytuowania względem rozwiązania optymalnego – może być związany tylko z jednym poziomem stabilności i może mieć tylko jeden zakres możliwych zmian wartości.

Miarą stabilności rozwiązania są zmiany wartości parametrów wejściowych, które zachowują przedstawione wcześniej warunki, tj. optymalność rozwiązania pierwotnego i odpowiedni (zawsze określony) rodzaj stabilności. Zmiany wartości parametrów określa się za pomocą długości zakresów i/lub granic zmienności. Długość zakresu zmienności określa, o ile może się zmniejszyć (długość dolnego zakresu) lub zwiększyć (długość górnego zakresu) przyjęta w zadaniu wartość parametru, aby stabilność rozwiązania była jeszcze zachowana. Wynika z tego, że długości zakresów zmienności nie mogą przyjmować wartości ujemnych. Granica zmienności określa, do jakiej wartości można zmniejszyć (granica dolna) lub zwiększyć (granica górna) dany parametr, aby stabilność rozwiązania była jeszcze zachowana. Możliwe jest również posługiwanie się pojęciem przedziału zmienności, określonym jako zbiór wartości nie mniejszych od granicy dolnej i jednocześnie nie większych od granicy górnej.

W ogólnym zapisie algebraicznym miary stabilności rozwiązania optymalnego można przedstawić w sposób następujący:

(1) $$GDp = p - ZDp,$$

(2) $$GGp = p + ZGp,$$

(3) $$PZp \in \langle GDp; GGp \rangle,$$

gdzie:

p – wartość parametru wejściowego;
GDp – dolna granica zmienności parametru wejściowego;
GGp – górna granica zmienności parametru wejściowego;
ZDp – długość dolnego zakresu zmian wartości parametru wejściowego;
ZGp – długość górnego zakresu zmian wartości parametru wejściowego;
PZp – przedział zmienności parametru wejściowego.

Cechą charakterystyczną ustalania miar stabilności jest to, że zakresy zmian wartości parametrów wejściowych są możliwe do obliczenia bez dodatkowego postępowania optymalizacyjnego opartego na algorytmie simpleksowym, a jedynie na podstawie analizy numerycznej pewnych fragmentów tablicy simpleksowej rozwiązania optymalnego.

---

[3]Niekiedy może to również doprowadzić do sprzeczności zadania.

Przyczyny trudności w analizie

Spośród trzech rodzajów parametrów występujących w wejściowym modelu programowania liniowego, tj. współczynników funkcji celu, wyrazów wolnych i współczynników technicznych, te ostatnie mają najskromniejszą literaturę dotyczącą ich zmienności. W literaturze poświęconej zastosowaniom analizy stabilności spotkać można jedynie próbę analizy współczynników technicznych przy niebazowych zmiennych decyzyjnych w ograniczeniach związanych z niebazowymi zmiennymi uzupełniającymi[4] (warunek napięty przy decyzji optymalnej) [6, s. 47–49]. Także literatura programowania liniowego wskazuje na trudności związane z analizą zmienności współczynników technicznych, a w szczególności znajdujących się przy bazowych zmiennych decyzyjnych w warunkach napiętych przy decyzji optymalnej. Trudności te wynikają stąd, że zmiana wartości jednego współczynnika technicznego może spowodować zmiany wartości wszystkich elementów tablicy simpleksowej rozwiązania optymalnego. Dodatkowym utrudnieniem analizy jest to, że skutki zmian wartości współczynników technicznych – w odróżnieniu od pozostałych parametrów – mają charakter nieliniowy [14, s. 260].

W literaturze programowania liniowego z lat 60. i 70. współczynniki techniczne przy bazowych zmiennych decyzyjnych w warunkach napiętych przy decyzji optymalnej były pomijane w analizie stabilności [3, s. 76; 15, s. 56]. W latach późniejszych spotkać się można z analizą zmian wartości więcej niż jednego współczynnika technicznego (kolumny lub wiersza) [14, s. 260–268; 19, s. 113––119], chociaż nie podaje się tam reguł dotyczących zmienności pojedynczego współczynnika technicznego, w odróżnieniu od pozostałych parametrów, czyli współczynników funkcji celu i wyrazów wolnych [14, s. 240; 19, s. 104-112]. Podobnie została potraktowana ta kwestia w monografii dotyczącej analizy pooptymalizacyjnej [7, s. 79 i 176 w porównaniu ze s. 258–272]. Ukazało się również opracowanie [8, s. 136–139], gdzie przedstawiono teoretyczne reguły zmian wartości współczynników technicznych dla zadań z minimalizowaną funkcją celu. Uzyskane wzory były na tyle skomplikowane (trójczłonowa formuła dla współczynników technicznych przy bazowych zmiennych decyzyjnych), że nie zilustrowano ich przykładem, który wykorzystano przy analizie zmian wartości współczynników funkcji celu i wyrazów wolnych [8, s. 134–135].

---

[4]Niebazowa zmienna uzupełniająca (dodatkowa) jest zawsze związana z warunkiem ograniczającym i służy do przekształcenia klasycznej lub mieszanej postaci zadania wejściowego do postaci standardowej. Zmienną uzupełniającą może być zmienna: swobodna niedoboru (znak relacji „≤"), swobodna nadmiaru (znak relacji „≥") lub sztuczna (znak relacji „=").

Założenia proponowanej metody analizy

Nie opracowano dotychczas prostych reguł rządzących zmianami wartości współczynników technicznych usytuowanych w zadaniu wejściowym przy bazowej zmiennej decyzyjnej w warunku napiętym przy decyzji optymalnej[5]. Modyfikacja wartości takiego współczynnika technicznego może spowodować zmiany wartości wszystkich danych znajdujących się w tablicy simpleksowej rozwiązania optymalnego[6]. Mimo iż znana jest w literaturze formuła Bodewiga dotycząca skutków zmian w macierzy odwrotnej bazy spowodowanych określoną zmianą elementów jednej linii (wiersza lub kolumny) w macierzy odwracanej [7, s. 257 i dalsze] oraz trójczłonowa formuła zmian wartości współczynnika technicznego dla zadań z minimalizowaną funkcją celu [8, s. 136–139], to jednak nie wyprowadzono dotychczas podobnie prostych formuł – umożliwiających na wzór pozostałych parametrów zadania wejściowego – określania długości zakresów/granic zmienności pojedynczych tego rodzaju współczynników technicznych.

Konfrontacja powyższych uwag z zakresem stabilności rozwiązania optymalnego prowadzi do wniosku, iż w przypadku współczynników technicznych znajdujących się w tablicy wejściowej przy bazowej zmiennej decyzyjnej w warunku napiętym przy decyzji optymalnej można mówić jedynie o najniższym poziomie stabilności, tj. stabilności bazy. Znaczy to, iż będzie należało poszukiwać takich zmian wartości współczynników technicznych, które nie naruszą zestawu zmiennych bazowych, natomiast mogą powodować zmiany ich wartości (nawet wszystkich zmiennych bazowych).

Z właściwości tablicy simpleksowej rozwiązania optymalnego [2, s. 40–41] wynika, iż zmiana wartości współczynnika technicznego wcale nie musi powodować utraty stabilności rozwiązania optymalnego, dla której na najniższym poziomie (stabilność bazy) wystarczy, aby żadna z wartości zmiennych bazowych, jak również żadna z wartości cen dualnych niebazowych zmiennych decyzyjnych/swobodnych nie zmieniła znaku. Ze względu na nieliniowe zmia-

---

[5]Oznacza to wykorzystanie ograniczenia przez decyzję optymalną na poziomie wartości wyrazu wolnego w tablicy wejściowej, a jednocześnie takie ograniczenie jest związane z którąś z niebazowych zmiennych uzupełniających (swobodną niedoboru, swobodną nadmiaru lub sztuczną).

[6]Jeżeli ulega zmianie wartość współczynnika technicznego (element macierzy odwracanej), wówczas może to spowodować w krańcowym przypadku zmianę (nieliniową) wartości wszystkich elementów (współczynników substytucji) macierzy odwrotnej bazy. Z kolei ta macierz jest źródłem zmian wartości w pozostałych elementach tablicy simpleksowej rozwiązania optymalnego.

ny wartości współczynników substytucji nie jest natomiast wiadome, gdzie najwcześniej nastąpi owa zmiana znaku (czy w kolumnie – cena dualna, czy w wierszu – zmienna bazowa).

W warunkach napiętych przy decyzji optymalnej (brak bazowych zmiennych swobodnych) istnieją zakresy możliwych zmian wartości wyrazów wolnych, które – podobnie jak w przypadku niniejszej analizy – powodują zmiany w wartościach zmiennych bazowych. Dlatego też patrząc przez pryzmat zmian wartości wyrazu wolnego w warunku napiętym przy decyzji optymalnej związanego z analizowanym współczynnikiem technicznym, należałoby się zastanowić, czy zmniejszanie/zwiększanie wartości wyrazu wolnego w ramach dozwolonego zakresu wpływa na wzrost/spadek wartości bazowej zmiennej decyzyjnej związanej z analizowanym współczynnikiem technicznym (poprzez analizę znaku wartości współczynnika substytucji umiejscowionego w wierszu bazowej zmiennej decyzyjnej i kolumnie niebazowej zmiennej uzupełniającej w tablicy simpleksowej rozwiązania optymalnego[7] i znaku relacji warunku ograniczającego, w którym znajduje się dany współczynnik techniczny, co jest tożsame z rodzajem niebazowej zmiennej uzupełniającej – swobodna niedoboru, swobodna nadmiaru, sztuczna). Przyczyną wzrostu/spadku wartości bazowej zmiennej decyzyjnej była oczywiście zmiana wartości wyrazu wolnego. Można stąd – przez analogię – domniemywać, że zamiast relacji:

zmiana wartości wyrazu wolnego & stała wartość współczynnika technicznego $\Rightarrow$ zmiana wartości bazowej zmiennej decyzyjnej,

będzie zachodziła również relacja:

zmiana wartości współczynnika technicznego & stała wartość wyrazu wolnego $\Rightarrow$ zmiana wartości bazowej zmiennej decyzyjnej.

W obu relacjach pozostaje niezmiennym zapisem „zmiana wartości bazowej zmiennej decyzyjnej". Stałość lub zmienność wartości współczynnika technicznego jest przyczyną stałości/zmienności wartości odpowiedniego współczynnika substytucji. Aby uzyskać rozmiar zmiany wartości współczynnika technicznego, należy poszukać wielkości, która mogłaby być wspólna dla

[7]Zmniejszanie wartości wyrazu wolnego będzie wpływało na wartość bazowej zmiennej decyzyjnej zgodnie z poniższym schematem (w – wzrost, s – spadek):

Znak relacji	Znak współczynnika substytucji	
	+	-
≤/=	s	w
≥	w	s

Natomiast zwiększanie wartości wyrazu wolnego będzie powodowało skutki przeciwne.

obu relacji, a jednocześnie byłaby wielkością znaną. Jedyną z nich jest długość zakresu zmienności wyrazu wolnego, z tym że zależność jest tu następująca:

zmiana wartości bazowej zmiennej decyzyjnej & stała wartość współczynnika substytucji $\Rightarrow$ długość zakresu zmienności wyrazu wolnego.

Na podstawie tej relacji można ustalić graniczne (minimalny lub maksymalny) poziomy wartości bazowej zmiennej decyzyjnej. Po tym ustaleniu można powrócić do wcześniejszej relacji:

zmiana wartości współczynnika technicznego & stała wartość wyrazu wolnego $\Rightarrow$ zmiana wartości bazowej zmiennej decyzyjnej,

i określić wartość zmiany współczynnika technicznego poprzez ustaloną wcześniej długość zakresu zmienności wyrazu wolnego. Następuje tutaj „przeniesienie" zmiany wartości wyrazu wolnego na zmianę wartości współczynnika technicznego przy tożsamych skutkach dla wartości zmiennych bazowych. Pamiętać przy tym należy, iż alternatywą do zwiększania wartości wyrazu wolnego jest zmniejszanie wartości współczynnika technicznego i na odwrót.

Wszystko to – co stwierdzono wyżej – będzie uzasadnione, jeśli w wyniku tych zmian (przede wszystkim wartości współczynników substytucji) któraś z wartości cen dualnych niebazowych zmiennych decyzyjnych/swobodnych nie zmieni znaku na przeciwny, czego przecież wykluczyć nie sposób. Dlatego też należałoby równocześnie analizować analogiczne relacje (oparte na tym samym współczynniku substytucji) dla zadania dualnego, co jest równoważne analizie długości zakresu zmienności współczynnika funkcji celu i wartości ceny dualnej niebazowej zmiennej uzupełniającej w zadaniu prymalnym. Znając relacje między wartościami cen dualnych (wskaźnikami optymalności) a wartościami zmiennych bazowych zadania prymalnego i dualnego można wyciągnąć wniosek, że jednocześnie trzeba badać w zadaniu prymalnym długość zakresu zmienności wyrazu wolnego i długość zakresu zmienności współczynnika funkcji celu oraz odpowiadającą im wartość bazowej zmiennej decyzyjnej i wartość ceny dualnej niebazowej zmiennej uzupełniającej. Mniejszy moduł z tych dwóch relacji wyznaczy długość zakresu zmienności analizowanego współczynnika technicznego.

Wyniki

Dolną granicę zmienności współczynnika technicznego znajdującego się przy bazowej zmiennej decyzyjnej „k" ($k = 1, 2, ..., n$) w warunku ograniczającym „w" ($w = 1, 2, ..., m$) – zgodnie ze wzorem (1) – oblicza się jako

(4) $$GDa_{wk} = a_{wk} - ZDa_{wk},$$

gdzie:
a_{wk} – wartość współczynnika technicznego w warunku ograniczającym „w" przy zmiennej decyzyjnej „k";
ZDa_{wk} – długość dolnego zakresu zmian wartości współczynnika technicznego przy zmiennej decyzyjnej „k" w warunku ograniczającym „w".

Górną granicę zmienności współczynnika technicznego znajdującego się przy bazowej zmiennej decyzyjnej „k" ($k = 1, 2, ..., n$) w warunku ograniczającym „w" ($w = 1, 2, ..., m$) – zgodnie ze wzorem (2) – oblicza się jako

(5) $$GGa_{wk} = a_{wk} + ZGa_{wk},$$

gdzie:
ZGa_{wk} – długość górnego zakresu zmian wartości współczynnika technicznego przy zmiennej decyzyjnej „k" w warunku ograniczającym „w".

Długość dolnego zakresu zmian wartości współczynnika technicznego przy bazowej zmiennej decyzyjnej w warunku napiętym przy decyzji optymalnej oblicza się ze wzorów:

1) gdy warunek jest o g r a n i c z o n y z g ó r y lub s z t y w n y

(6) $$ZDa_{wk} = min\left[\left(\frac{x_k}{ZGb_w} + r_{hv}\right)^{-1};\ \left(\frac{|p_w|}{ZGc_k} + r_{hv}\right)^{-1}\right],$$

gdzie:
x_k – wartość zmiennej decyzyjnej „k";
h – numer wiersza w tablicy simpleksowej rozwiązania optymalnego, w którym znajduje się bazowa zmienna decyzyjna „k";
v – numer kolumny w tablicy simpleksowej rozwiązania optymalnego, w której znajduje się niebazowa zmienna uzupełniająca związana z warunkiem ograniczającym „w";
r_{hv} – wartość współczynnika substytucji znajdującego się na przecięciu wiersza „h" i kolumny „v" w tablicy simpleksowej rozwiązania optymalnego;
p_w – wartość ceny dualnej niebazowej zmiennej uzupełniającej związanej z warunkiem ograniczającym „w";
ZGb_w – długość górnego zakresu zmian wartości wyrazu wolnego w warunku ograniczającym „w"

$$ZGb_w = \begin{cases} min\left(\dfrac{-x_i}{r_{iv}}\right) : r_{iv} < 0, \quad dla \ i = 1,2,...,m, \\ \infty, \ gdy \ nie \ ma \ r_{iv} < 0, \quad dla \ i = 1,2,...,m; \end{cases}$$

ZGc_k – długość górnego zakresu zmian wartości współczynnika funkcji celu przy zmiennej decyzyjnej „k"

a) maksymalizacja funkcji celu

$$ZGc_k = \begin{cases} min\left|\dfrac{p_j}{r_{hj}}\right| : r_{hj} < 0, \quad dla \ j = 1,2,...,n-s, \\ \infty, \ gdy \ nie \ ma \ r_{hj} < 0, \quad dla \ j = 1,2,...,n-s; \end{cases}$$

b) minimalizacja funkcji celu

$$ZGc_k = \begin{cases} min\left|\dfrac{p_j}{r_{hj}}\right| : r_{hj} > 0, \quad dla \ j = 1,2,...,n-s, \\ \infty, \ gdy \ nie \ ma \ r_{hj} > 0, \quad dla \ j = 1,2,...,n-s; \end{cases}$$

s – liczba warunków sztywnych (znak relacji „=");

2) gdy warunek jest ograniczony z dołu

(7) $$ZDa_{wk} = min\left[\left(\frac{x_k}{ZGb_w} - r_{hv}\right)^{-1}; \ \left(\frac{|p_w|}{ZGc_k} - r_{hv}\right)^{-1}\right],$$

gdzie:

$$ZGb_w = \begin{cases} min\left(\dfrac{x_i}{r_{iv}}\right) : r_{iv} > 0, \quad dla \ i = 1,2,...,m, \\ \infty, \ gdy \ nie \ ma \ r_{iv} > 0, \quad dla \ i = 1,2,...,m. \end{cases}$$

Długość górnego zakresu zmian wartości współczynnika technicznego przy bazowej zmiennej decyzyjnej w warunku napiętym przy decyzji optymalnej oblicza się ze wzorów:

1) gdy warunek jest o g r a n i c z o n y z g ó r y lub s z t y w n y

$$ZGa_{wk} = min\left[\left(\frac{x_k}{ZDb_w} - r_{hv}\right)^{-1};\ \left(\frac{|p_w|}{ZDc_k} - r_{hv}\right)^{-1}\right], \tag{8}$$

gdzie:
ZDb_w – długość dolnego zakresu zmian wartości wyrazu wolnego w warunku ograniczającym „w"

$$ZDb_w = \begin{cases} min\left(\frac{x_i}{r_{iv}}\right) : r_{iv} > 0, \quad dla\ i = 1,2,...,m, \\ \infty,\ gdy\ nie\ ma\ r_{iv} > 0, \quad dla\ i = 1,2,...,m; \end{cases}$$

ZDc_k – długość dolnego zakresu zmian wartości współczynnika funkcji celu przy zmiennej decyzyjnej „k"

a) m a k s y m a l i z a c j a f u n k c j i c e l u

$$ZDc_k = \begin{cases} min\left|\frac{p_j}{r_{hj}}\right| : r_{hj} > 0, \quad dla\ j = 1,2,...,n-s, \\ \infty,\ gdy\ nie\ ma\ r_{hj} > 0, \quad dla\ j = 1,2,...,n-s; \end{cases}$$

b) m i n i m a l i z a c j a f u n k c j i c e l u

$$ZDc_k = \begin{cases} min\left|\frac{p_j}{r_{hj}}\right| : r_{hj} < 0, \quad dla\ j = 1,2,...,n-s, \\ \infty,\ gdy\ nie\ ma\ r_{hj} < 0, \quad dla\ j = 1,2,...,n-s; \end{cases}$$

2) gdy warunek jest o g r a n i c z o n y z d o ł u

$$(9) \qquad ZGa_{wk} = min\left[\left(\frac{x_k}{ZDb_w} + r_{hv}\right)^{-1};\ \left(\frac{|p_w|}{ZDc_k} + r_{hv}\right)^{-1}\right],$$

gdzie:

$$ZDb_w = \begin{cases} min\left(\dfrac{-x_i}{r_{iv}}\right) : r_{iv} < 0, \quad dla\ i = 1,2,...,m, \\ \infty,\ gdy\ nie\ ma\ r_{iv} < 0, \quad dla\ i = 1,2,...,m. \end{cases}$$

Jeśli dzielnikiem w którejś z podstaw potęgowania wzorów (6)–(9) jest nieskończoność (∞), to przyjmuje się, iż wynikiem danego ilorazu jest zero (0), a jeśli któraś z podstaw potęgowania nie jest dodatnia, to przyjmuje się, iż wynikiem danego potęgowania jest nieskończoność (∞).

Literatura

1. BUGA J., Wybrane problemy zastosowań metod programowania matematycznego, [w:] W. Sadowski (red.), *Elementy ekonometrii i programowania matematycznego*, PWN, Warszawa 1980.
2. BUSŁOWSKI A., *Stabilność rozwiązania optymalnego zadania programowania liniowego*, Wydawnictwo Uniwersytetu w Białymstoku, Białystok 2000.
3. CYBURA A., JUCHNOWSKI M., Problemy parametryzacji współczynników funkcji celu w optymalizacyjnych rolniczych modelach liniowych, *Zagadnienia Ekonomiki Rolnej*, 1979, nr 4.
4. *Encyklopedia ekonomiczno-rolnicza*, PWRiL, Warszawa 1984.
5. GAJEWSKI J., *Sposoby aktualizacji optymalnych planów w przedsiębiorstwach rolniczych*, IRWiR PAN, Warszawa 1974.
6. GAJEWSKI J., ANDRYCHOWICZ B., *Zmiany warunków gospodarowania a planowanie w przedsiębiorstwie rolniczym*, PWRiL, Warszawa 1979.
7. GAL T., *Postoptimal Analyses. Parametric Programming, and Related Topics*, McGraw-Hill International Book Company 1979.
8. GASS S.I., *Programowanie liniowe. Metody i zastosowania*, PWN, Warszawa 1973.

9. GRUDA M., Elementy analizy stabilności modeli programowania liniowego z przykładami zastosowań, [w:] J. Gajewski (red.), *Informacja a zarządzanie w PGR*, PWRiL, Warszawa 1983.
10. HEADY E.O., CANDLER W., Metody programowania liniowego, [w:] K. Rey i A. Woś (red.), *Metody matematyczne w ekonomice i planowaniu rolnictwa*, PWRiL, Warszawa 1965.
11. KOŁATKOWSKI D., Zagadnienia pooptymalizacyjne, [w:] S. Dorosiewicz, M. Gruszczyński, D. Kołatkowski, T. Kuszewski, M. Podgórska, E. Syczewska, *Ekonometria*, Oficyna Wydawnicza SGH, Warszawa 1996.
12. MARSZAŁKOWICZ T., *Metody programowania optymalnego w rolnictwie*, PWE, Warszawa 1986.
13. NYKOWSKI I., Programowanie liniowe, [w:] W. Sadowski (red.), *Elementy ekonometrii i programowania matematycznego*, PWN, Warszawa 1980.
14. NYKOWSKI I., *Programowanie liniowe*, PWE, Warszawa 1984.
15. ORKISZ T., Analiza stabilności optymalnego rozwiązania zadania programowania liniowego (metodą simpleks), *Przegląd Statystyczny*, 1964, tom XI, zeszyt 1.
16. PIETRASZEWSKI A., WAGNER W., WYSOCKI F., *Podstawy agroekonometrii*, Wydawnictwo AR, Poznań 1989.
17. PODKAMINER L., Wpływ zmian parametrów techniczno-ekonomicznych na wielkość modelowego optimum gospodarstwa rolnego, *Zagadnienia Ekonomiki Rolnej*, 1973, nr 2.
18. RACZKOWSKA U., BUSŁOWSKI A., Analiza zmian wielkości zmiennych decyzyjnych w rozwiązaniu optymalnym zadania programowania liniowego, *Optimum – Studia Ekonomiczne*, 1998, nr 1.
19. ROGALSKA D. (red.), *Programowanie liniowe. Algorytmy i zadania*, Uniwersytet Łódzki, Łódź 1983.
20. WOŚ A., *Rachunek ekonomiczny w rolnictwie*, PWRiL, Warszawa 1966.

Technical coefficients and stability of optimal solution of linear programming problem

Abstract

Profile of coefficients in constrain conditions of linear programming problem described in farm example can conclude that there are significant number of fixed coefficients without possibility of changeability analysis. It is important that this group of coefficients is easy detected because their value is equal to 0 or ±1. They were called logical coefficient as opposed from technical coefficients. In the paper formulas describing changeability of technical coefficients of real basic variables with co-called tight conditions in optimal decision are shown. These formulas was based on linear effects of changing cost and right-hand side coefficients in spite of nonlinear dependence between changing technical coefficients and basis variables. Thanks to such solution it is easy to obtain stability measure in case of changing technical coefficients.

Aneta Becker
Katedra Statystyki Matematycznej AR w Szczecinie

Metody mierzenia efektywności zarządzania w przedsiębiorstwie

Wstęp

Efektywność jest jednym z najczęściej używanych pojęć w naukach ekonomicznych. Dotyczy stosunku między wartością poniesionych nakładów a wartością efektów uzyskanych dzięki tym nakładom. W szerszym ujęciu efektywność ekonomiczna oznacza uzyskiwanie najlepszych rezultatów w produkcji bądź dystrybucji towarów i usług po najniższych kosztach. Występująca pod pojęciem sprawności ekonomicznej oznacza zdolność jednostki gospodarczej do wytwarzania w danym czasie za pomocą określonych środków dóbr i usług zapewniających zaspokojenie potrzeb odbiorcy. Efektywność określoną za pomocą relacji efektów i nakładów można wykorzystać w praktyce gospodarowania do:

- analizy retrospektywnej (ex post) efektywności wykorzystywania dotychczas posiadanych zasobów,
- analizy prospektywnej (ex ante) jako narzędzia pomocniczego przy podejmowaniu decyzji dotyczących przyszłości.

Efektywność można traktować jako wynik połączenia dwóch komponentów, a mianowicie:

- skuteczności działania określonej jako stopień zachodzenia na siebie deklarowanego celu działania i uzyskanego w wyniku tego działania ostatecznego rezultatu,
- sprawności funkcjonowania jako relacji między uzyskiwanymi rezultatami działania (efektami) a poniesionymi nakładami (zaangażowanymi środkami).

Określeniami zbliżonymi do pojęcia efektywności są gospodarność i racjonalność. Za racjonalne działanie uważa się takie, które prowadzone jest w najlepszej wierze, zgodnie ze zdrowym rozsądkiem i wiedzą realizujących je osób. Natomiast gospodarność to ogólnie lepsze, w porównaniu do stanu przeciętnego, wyniki w zakresie realizacji celów działalności jednostki gospodaru-

jącej. Przez gospodarność rozumie się także ogólnie lepszy, w porównaniu do przeciętnego, stopień wykorzystania istniejących w dyspozycji przedsiębiorstwa środków rzeczowych i kadr, w wyniku czego koszty własne produkcji przedsiębiorstwa spadają poniżej oczekiwanego, przeciętnego przy danej skali produkcji i przy danych warunkach technicznych i organizacyjnych poziomu. Jeżeli analizę gospodarności przeprowadza się z punktu widzenia rozmiarów produkcji, to przez pojęcie gospodarności rozumie się ogólnie lepsze, w porównaniu do stanu przeciętnego, wyniki przedsiębiorstwa w zakresie realizacji zmiennych będących uściśleniem zasady racjonalnego gospodarowania w danych warunkach techniczno-organizacyjnych i ekonomicznych.

W analizie gospodarności dokonuje się konfrontacji efektów działania pojedynczej lub wielu organizacji w kolejnych okresach z pewnym stanem uznanym za przeciętny. Analiza gospodarności jest jedną z funkcji procesu zarządzania. Należy do kontrolnej funkcji zarządzania i zajmuje się problematyką kryteriów oceny oraz metod oceny działalności organizacji.

Ważnym pojęciem, które należy zdefiniować mówiąc o efektywności gospodarowania, jest efektywność postępu w przedsiębiorstwie. Ogólnie postęp jest to proces zmian przechodzących od stanów mniej doskonałych do doskonalszych. Inna definicja mówi, że przez postęp rozumie się pozytywny rozwój ilościowo-jakościowy systemu w czasie, tj. zmiany stanów (warunków) działalności przedsiębiorstwa[1]. Można wyróżnić dwa główne rodzaje postępów, mianowicie: postęp techniczny, polegający na doskonaleniu środków produkcji, metod wytwarzania, zastępowaniu pracy żywej przez środki techniczne, oraz postęp ekonomiczny, który ma miejsce, gdy w przedsiębiorstwie zachodzą korzystne zmiany polegające na wzroście gospodarności i lepszym wykorzystaniu posiadanych czynników produkcji, poprawie efektywności gospodarowania[2]. Innymi, często wymienianymi przy okazji mierzenia efektów postępu w przedsiębiorstwie postępami są: organizacyjny, czyli postęp przyczyniający się do lepszej realizacji celów przedsiębiorstwa, oraz kadrowy, związany z zatrudnieniem, kwalifikacjami pracowników itd. W literaturze można spotkać propozycje trzech sposobów pomiaru efektów postępu w przedsiębiorstwie:

- budowę przyczynowo-skutkowych modeli produkcji, gdzie do zbioru zmiennych objaśniających (oprócz czynników produkcji) wprowadza się zmienne mierzące poziom techniczny, organizacyjny, kadrowy i ekonomiczny,

---

[1]Por. Hozer J., Zawadzki J., *Proces ekonometrycznego modelowania*, US, Rozprawy i Studia, T./CXXV/51, Szczecin 1990, s. 118.

[2]Por. *Słownik ekonomiczny*, Wyd. ZNICZ, Szczecin 1994, s. 149.

- wprowadzenie do funkcji produkcji i funkcji kosztów jednostkowych zmiennej czasowej lub jej funkcji,
- wykorzystanie reszt odpowiednio wyspecyfikowanych modeli produkcji i kosztów jednostkowych.

Przedstawione powyżej podejście pozwala zmierzyć i oddzielić efekty obiektywnego (danego z zewnątrz) postępu od efektów postępu subiektywnego (zależnego od przedsiębiorstwa)[3].

Miary analizy ekonomicznej w ocenie efektywności gospodarowania

Głównym przedmiotem analiz oceny efektywności ekonomicznej przedsiębiorstwa jest sposób określenia mierników ilościowych, służących do ich klasyfikacji. W naukach ekonomicznych miernikami nazywa się wzory algebraiczne. Dokonywanie pomiarów za pomocą mierników ekonomicznych służy do takiego przyporządkowania liczb wartościom cech przedmiotów, aby z relacji zachodzących między tymi liczbami móc wnioskować o zachodzeniu izomorficznych relacji między cechami przedmiotów, którym liczby te przyporządkowano[4]. Mierniki oceny gospodarności służą do mierzenia stopni realizacji zasady racjonalnego działania w taki sposób, aby uzyskać dodatnią różnicę między dochodami z działalności gospodarczej a kosztami poniesionymi na tę działalność. W ekonomii nazwy miernik używa się zamiennie z nazwą wskaźnik (ma się na myśli wskaźniki ilościowe), które stanowią podstawę do twierdzenia, o ile pewne działania gospodarcze są lepsze albo gorsze, bardziej lub mniej korzystne od innych działań. Wszystkie mierniki oceny działalności przedsiębiorstwa można podzielić na następujące grupy:

- mierniki rozmiarów produkcji – wielkość i ilość wykonanych wyrobów i usług (np. produkcja globalna, produkcja towarowa, sprzedaż),
- mierniki jakości pracy (np. zysk, akumulacja, wydajność pracy, koszty jednostkowe),
- mierniki stanu techniczno-organizacyjnego i ekonomicznego przedsiębiorstw (np. rozmiary, struktura i wykorzystanie czynników produkcji, wielkość inwestycji i ich struktura)[5].

---

[3]Por. Hozer J., Zawadzki J., op. cit., s. 120–121.

[4]Stachak S., *Ekonomika agrofirmy*, PWN, Warszawa 1998, s. 343.

[5]Por. Urbańczyk E., *Metody analizy ekonomicznej efektywności majątku trwałego w przemyśle*, Prace Naukowe Politechniki Szczecińskiej nr 300, Szczecin 1985, s. 33.

Przy czym jeśli chodzi o ocenę gospodarowania, to należy korzystać z mierników sklasyfikowanych w drugiej i trzeciej grupie.

Efektywność ekonomiczną przedsiębiorstwa można mierzyć, konfrontując różne postaci otrzymywanych wyników z nakładami poniesionymi na ich rzecz. W związku z tym wyróżnia się trzy formuły rachunku efektywności ekonomicznej, różniące się zakresem ujmowanych nakładów[6].

- Pierwsza obejmuje wyniki (produkcyjne, usługowe, finansowe itp.) przedsiębiorstwa, które odnoszone są do bieżących nakładów pracy żywej i uprzedmiotowionej; w rezultacie otrzymuje się *nakładowe mierniki efektywności ekonomicznej przedsiębiorstwa* w wyrazie strumieniowym.

Nakładowe mierniki efektywności ekonomicznej przedsiębiorstwa odzwierciedlają zmiany efektywności, które są spowodowane poprawą wyników ekonomicznych oraz oszczędnością bieżących nakładów pracy żywej i uprzedmiotowionej. Nakładowy miernik można zapisać w następujący sposób:

$$E = \frac{P}{\frac{t_1 + t_2}{r}} \tag{1}$$

gdzie:

E – miernik efektywności ekonomicznej przedsiębiorstwa,

P – miernik produkcji brutto,

t_1 – nakłady pracy żywej wyrażone w jednostkach czasu pracy lub liczbie pracowników,

t_2 – nakłady pracy uprzedmiotowionej,

r – produkcja czysta przypadająca na jednostkę nakładu pracy żywej.

- Druga uwzględnia udział w powstawaniu wyników określonych zasobów środków trwałych jako skapitalizowanych nakładów inwestycyjnych oraz pracy ludzkiej, które są rezultatem nakładów społecznych poniesionych na wytworzenie potencjału kwalifikacyjnego zatrudnienia; w rezultacie otrzymuje się *zasobowe mierniki efektywności ekonomicznej przedsiębiorstwa.*

W tej grupie mierników ważny jest sposób dodawania poszczególnych elementów: zasobów, czynników produkcji i zatrudnienia. Nakłady jednorazowe są porównywane z produkcją netto, a nie brutto, z powodu powtórnego rachunku elementów składowych. Wyjściowy zasobowy miernik efektywności ekonomicznej przedsiębiorstwa ma postać:

---

[6]Por. Czechowski L., *Wielowymiarowa ocena efektywności ekonomicznej przedsiębiorstwa przemysłowego*, Wyd. Uniwersytetu Gdańskiego, Gdańsk 1997, s. 80.

$$E = \frac{P}{t_1 + t_2} \tag{2}$$

gdzie:
E – miernik efektywności ekonomicznej przedsiębiorstwa,
P – miernik produkcji netto,
t_1 – liczba zatrudnionych pracowników sprowadzona do współmierności z funduszami produkcyjnymi (kapitałami),
t_2 – środki produkcji sprowadzone do współmierności z zatrudnieniem.

- Trzecia jest wynikiem zastosowania obu wymienionych powyżej formuł, czego rezultatem są mieszane współczynniki, a także wskaźniki efektywności ekonomicznej.

Formuła ta jest relacją wyników produkcyjnych do sumy nakładów, które są kombinacją bieżących nakładów pracy żywej (czasu roboczego) i pracy uprzedmiotowionej (przedmiotów pracy i amortyzacji) oraz nakładów przeszłych okresów (jednorazowych), zainwestowanych w przeszłości na utworzenie i rozwój czynników wytwórczych, czyli zasobów majątku produkcyjnego i wykwalifikowanej pracy żywej. Podstawową formę mieszanego (nakładowo-zasobowego) miernika efektywności ekonomicznej przedsiębiorstwa można zapisać w następujący sposób:

$$E = \frac{P}{K + N \cdot T} \tag{3}$$

gdzie:
E – miernik efektywności ekonomicznej przedsiębiorstwa,
P – produkcja czysta w cenach stałych lub produkcja globalna w cenach bieżących,
K – koszty własne produkcji,
N – normatyw efektywności kapitału,
T – kapitał trwały i obrotowy (bez funduszu płac) w cenach porównywalnych.

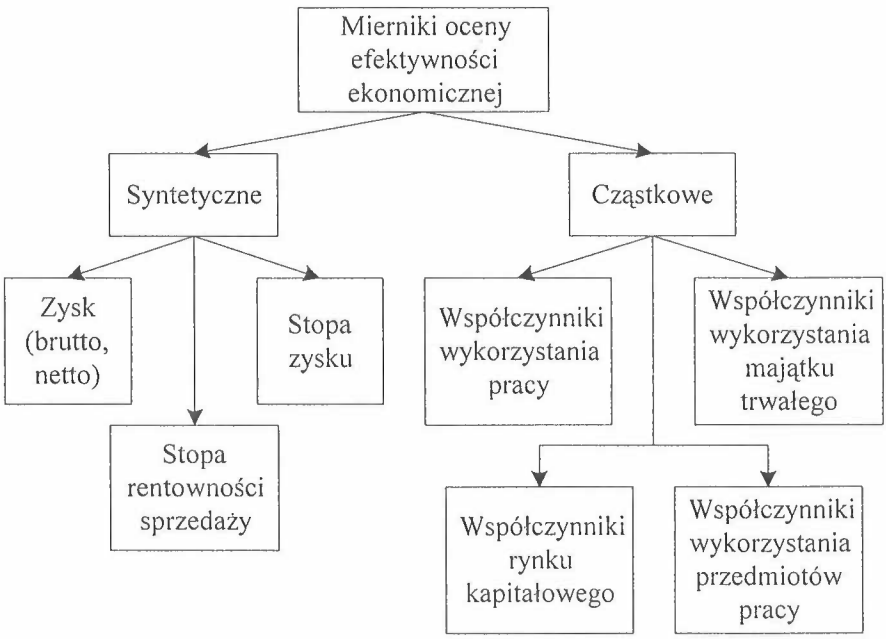
Grupy mierników zaprezentowane powyżej, stosowane w przeszłości, obarczone są pewnymi wadami, do których można zaliczyć to, że:

- nie miały charakteru obiektywnego,
- w pewnej części ustalane były przez ówczesnych decydentów w oderwaniu od zasad rachunku ekonomicznego,
- były przedmiotem bardziej rozważań teoretycznych niż możliwości zastosowania w praktyce gospodarczej.

Istnieje wiele mierników, które są stosowane w ocenie efektywności ekonomicznej przedsiębiorstwa, co powoduje problemy z ich klasyfikacją. Przyj-

mując pojemność (liczbę zdarzeń gospodarczych) za kryterium podziału, mierniki te dzielą się na (schemat 1):

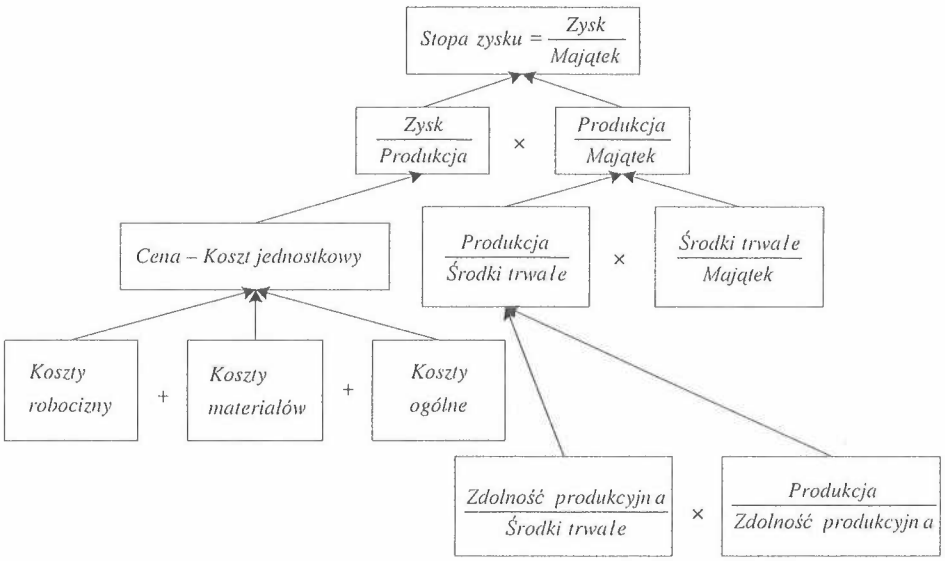
- syntetyczne, zawierające w kompleksowy sposób wszystkie czynniki charakteryzujące efektywność badanego zamierzenia oraz odznaczające się dużą pojemnością; otrzymuje się je w wyniku podzielenia przez siebie całości uzyskanych wyników i całości poniesionych nakładów, co oznacza, że są one prawie zawsze współczynnikami i wskaźnikami typu ilorazowego, wyjątek stanowi absolutna kwota zysku (różnica między przychodami ze sprzedaży a kosztami ich uzyskania);
- cząstkowe, stosowane w celach analitycznych jako instrument prostej i jednoznacznej oceny procesu gospodarowania poszczególnymi czynnikami produkcji; do podstawowych mierników cząstkowych można zaliczyć: współczynniki produktywności czynników wytwórczych (np. wydajność pracy, produktywność środków trwałych itp.) oraz współczynniki nakładochłonności produkcji (np. pracochłonność produkcji, majątkochłonność produkcji itp.)[7].



Schemat 1
Podział mierników oceny efektywności ekonomicznej przedsiębiorstw
Źródło: Opracowanie własne.

[7] Por. Czechowski L., op. cit., s. 88 i dalsze.

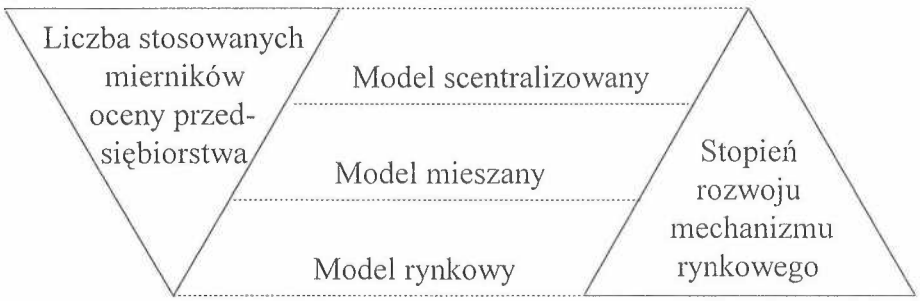
Między syntetycznymi a cząstkowymi miernikami efektywności ekonomicznej przedsiębiorstwa występuje zależność polegająca na tym, że mierniki cząstkowe mogą uzupełniać pomiar zjawisk, które nie są objęte miernikiem syntetycznym lub objąć swoim zasięgiem cały miernik syntetyczny, dzieląc go na poszczególne części, będące powieleniem zjawisk prezentowanych w sposób zbiorczy przez miernik syntetyczny. Przyjmuje się założenie, że stosując zarówno mierniki syntetyczne, jak i cząstkowe, te pierwsze odgrywają nadrzędną rolę, natomiast mierniki cząstkowe pełnią funkcję uzupełniającą i służą rozwinięciu treści miernika syntetycznego.



Schemat 2
Powiązania mierników cząstkowych z syntetycznym miernikiem efektywności ekonomicznej (stopą zysku)
Źródło: Czechowski L.., op. cit., s. 94.

Liczba mierników zależy od systemu zarządzania gospodarką. Im bardziej państwo ingeruje w procesy gospodarcze, stosując przy tym dużą liczbę mierników oceny działalności, tym mniejszy jest stopień rozwoju mechanizmu rynkowego[8].

[8] Por. Czechowski L., op. cit., s. 95.



Rysunek 1
Zależność między liczbą stosowanych mierników a stopniem rozwoju mechanizmu rynkowego
Źródło: Czechowski L.., op. cit., s. 95.

Zmiana systemu gospodarowania, jaka dokonała się w Polsce, pociąga za sobą zmianę zachowań przedsiębiorstw. Uzyskana samodzielność firm połączona jest z odpowiedzialnością za ich funkcjonowanie oraz rozwój w warunkach konkurencji. Strategia przedsiębiorstwa oraz jej realizacja gwarantująca sukces wywołuje zapotrzebowania na analizę finansową nie tylko typu ex post (retrospektywną, przeszłościową), ale i ex ante (prospektywną, przyszłościową)[9]. Analiza jest instrumentem badawczym, który służy do poznania rzeczywistości gospodarczej (przebiegu i rezultatów procesów gospodarczych). Natomiast analiza ekonomiczna to sposób badania procesów ekonomicznych za pomocą rozkładania ich na poszczególne elementy i rozpatrywanie wzajemnych związków między tymi elementami i procesami[10]. Analiza finansowa jest tą częścią analizy ekonomicznej, która obejmuje zagadnienia związane z całą działalnością przedsiębiorstwa, czyli sytuację majątkową, sytuację finansową (położenie i pewność finansowa), analizę wzrostu, analizę pozycji finansowej przedsiębiorstwa, wynik finansowy, a co za tym idzie ocenę efektywności gospodarowania, czyli rentowności, a także koszty i przychody ze sprzedaży[11].

---

[9]Por. Waśniewski T., Skoczylas W., *Zasady analizy finansowej w praktyce. Przykłady i zadania*, Fundacja Rozwoju Rachunkowości w Polsce, Warszawa 1994, s. 9.

[10]*Słownik ekonomiczny*, op. cit, s. 11.

[11]Por. Waśniewski T., Skoczylas W., op. cit, s. 9.

Do oceny całościowej kondycji finansowej służą wskaźniki finansowe, które prezentują relacje występujące pomiędzy wybranymi elementami sprawozdań finansowych. Ocena taka obejmuje:

a) badanie sytuacji majątkowej, której przedmiotem oceny są:
   - wielkość, struktura i dynamika majątku, czyli badanie zmian zachodzących w relacji między majątkiem trwałym a obrotowym oraz zmian w strukturze poszczególnych składników zarówno jednej, jak i drugiej zbiorowości; w ocenie sytuacji majątkowej duże znaczenie poznawcze przypisuje się następującym wskaźnikom:
     – wskaźnik struktury majątku trwałego,
     – wskaźnik struktury majątku obrotowego,
     – podstawowy wskaźnik struktury aktywów,

$$S_A = \frac{M_T}{M_O} \tag{4}$$

gdzie:
S_A – wskaźnik struktury aktywów,
M_T – majątek trwały,
M_O – majątek obrotowy;

   - wykorzystanie majątku trwałego i obrotowego (wskaźniki produktywności-obrotowości); związane jest z tym badanie posiadanych zdolności produkcyjnych, dokonywane za pomocą następujących wskaźników:
     – produktywności majątku ogółem,
     – produktywności rzeczowego i niematerialnego majątku trwałego,
     – obrotowości zapasów,
     – obrotowości należności;
   - polityka inwestycyjna oraz w zakresie odpisów amortyzacyjnych, która jest zagadnieniem ważnym, dostarczającym informacji odnośnie przeciętnego wieku, kształtowania się oraz stopnia modernizacji rzeczowego i niematerialnego majątku trwałego; charakteryzowana przez wskaźniki:
     – zużycia rzeczowego i niematerialnego majątku trwałego,
     – inwestycji,
     – intensywności odpisów amortyzacyjnych,
     – pokrycia nakładów inwestycyjnych z odpisów amortyzacyjnych;

b) badanie sytuacji kapitałowej; w analizie tej dąży się do badania tendencji rozwoju struktury pasywów, wyrażającej źródła finansowania firmy, a także powstałych w tym zakresie zmian; obejmuje ono ocenę:

- wielkości, struktury i dynamiki pasywów bilansu, której podstawą są odchylenia bezwzględne, wskaźniki tempa wzrostu oraz wskaźniki struktury ujęte w bilansie analitycznym;
- zadłużenia oraz niezależności finansowej, którą można scharakteryzować za pomocą następujących wskaźników:
  - wskaźnika zadłużenia,
  - podstawowego wskaźnika struktury pasywów,

$$S_P = \frac{K_W}{K_O} \tag{5}$$

gdzie:
S_P – wskaźnik struktury pasywów,
K_W – kapitał własny,
K_O – kapitał obcy;

  - wskaźnika natężenia rezerw,
  - wskaźnika samofinansowania,
  - wskaźnika zdolności-pewności kredytowej;
- efektywności wykorzystania kapitałów; dotyczy on oceny obrotowości zobowiązań i można wyróżnić tu:
  - wskaźnik szybkości obrotu kapitału całkowitego (własnego, obcego),
  - wskaźnik obrotowości zobowiązań;

c) badanie sytuacji kapitałowo-majątkowej, dzięki któremu uzyskuje się rzeczywisty obraz sytuacji ekonomiczno-finansowej przedsiębiorstwa; przedmiotem tej analizy są poniższe zagadnienia:
   - sposób finansowania majątku trwałego; obejmuje on ocenę wskaźników pokrycia I, II i III stopnia,

$$S_{API} = \frac{K_W}{M_T} \qquad S_{APII} = \frac{K_W + D_D}{M_T} \qquad S_{APIII} = \frac{K_W + D_D}{M_T + S_O} \tag{6}$$

gdzie:
S_{AP} – wskaźnik struktury kapitałowo-majątkowej (pokrycia I i kolejno II, III stopnia),
K_W – kapitał własny,
M_T – majątek trwały,
D_D – zobowiązania długoterminowe,
S_O – środki obrotowe (związane długookresowo);

- analiza płynności finansowej; *płynność* jest to zdolność przedsiębiorstwa do wywiązania się z bieżących zobowiązań; w analizie tej wykorzystuje się w niej:
  – wskaźniki płynności finansowej I stopnia,

$$WP_N = \frac{SP + KP}{P_B} \tag{7}$$

gdzie:
WP_N – wskaźnik płynności natychmiastowy,
SP – środki pieniężne,
KP – krótkoterminowe papiery wartościowe,
P_B – pasywa bieżące;

  – wskaźniki płynności finansowej II stopnia (Quick ratio),

$$WP_S = \frac{A_B - Z_A}{P_B} \tag{8}$$

gdzie:
WP_S – wskaźnik płynności szybki,
A_B – aktywa bieżące,
P_B – pasywa bieżące,
Z_A – zapasy;

  – wskaźniki płynności finansowej III stopnia (Current ratio),

$$WP_B = \frac{A_B}{P_B} \tag{9}$$

gdzie:
WP_B – wskaźnik bieżącej płynności finansowej,
A_B – aktywa bieżące,
P_B – pasywa bieżące;

- wielkość zmiany kapitału pracującego, gdzie *kapitał pracujący* jest to różnica między majątkiem obrotowym (krótkoterminowym) a zobowiązaniami bieżącymi, czyli wyraża wielkość kapitałów stałych (własnych i długoterminowych obcych) finansujących majątek obrotowy;
- badanie zapotrzebowania na kapitał pracujący;

d) badanie wyniku finansowego; *wynik finansowy* jest syntetyczną miarą efektów finansowych osiąganych przez przedsiębiorstwo w ramach prowa-

dzonej przez nie działalności gospodarczej; jest podstawowym miernikiem oceny działalności jednostki gospodarczej, podstawą oceny gospodarności, a także źródłem finansowania rozwoju przedsiębiorstwa przez reprodukcję kapitałów; wynik finansowy może być wielkością dodatnią (zysk) lub ujemną (strata); przedstawia się go ogólnie jako różnicę pomiędzy przychodami i kosztami ich uzyskania; związane są z nim dwa pojęcia, od których zależy sytuacja gospodarcza przedsiębiorstwa, a mianowicie rentowność i ryzyko; głównymi celami działalności przedsiębiorstwa są: maksymalizacja zysku i udziału w rynku, przetrwanie, bezpieczeństwo oraz osiągnięcie zadowalającego poziomu zysku, który jest głównym czynnikiem rzutującym na poziom rentowności i stanowi źródło zasilania w kapitał własny, niezbędny dla dalszego rozwoju przedsiębiorstwa; podstawowym źródłem analizy wyniku finansowego jest jego syntetyczne zestawienie, natomiast zasadnicze wnioski otrzymuje się w toku analizy przyczynowej, w której zakłada się, że na wynik finansowy mają wpływ przychody ze sprzedaży, koszty własne sprzedaży, saldo operacji finansowych, saldo strat i zysków nadzwyczajnych, podatek dochodowy; przy podejmowaniu decyzji krótkoterminowych ważnym instrumentem jest analiza progu rentowności, będącego wielkością, przy której przychody ze sprzedaży całkowicie pokrywają wysokość poniesionych kosztów;

e) badanie rentowności (ze szczególnym uwzględnieniem rentowności kapitału własnego); *rentowność* jest miarą efektywności gospodarowania, za pomocą której określa się opłacalność działań gospodarczych; osiąganie przez przedsiębiorstwo zysku informuje, że jest ono rentowne, natomiast w sytuacji odwrotnej (straty) firma jest nierentowna; można wyróżnić dwie grupy wskaźników rentowności:
   - wskaźniki opłacalności zaangażowanego kapitału, do grupy tej zalicza się wskaźniki przedstawiające iloraz zysku (wybraną wielkości wyniku) do kapitału (wybrany rodzaj kapitału);
   - wskaźniki efektywności zużycia czynników produkcji, którą można scharakteryzować za pomocą wskaźników stanowiących relację zysku (wybrana wielkości wyniku) do określonego rodzaju obrotu (koszt własny sprzedaży, przychód ze sprzedaży); do najczęściej używanych zalicza się następujące wskaźniki:
     – rentowności sprzedaży – marża zysku; może być liczony jako rentowność brutto i netto,

$$R_{SB} = \frac{Z_B}{P_N} \quad \text{lub} \quad R_{SN} = \frac{Z_N}{P_N} \tag{10}$$

gdzie:
R_{SB} – wskaźnik rentowności brutto,
Z_B – zysk brutto,
P_N – przychody ze sprzedaży netto (przychody bez podatku VAT),
R_{SN} – wskaźnik rentowności netto,
Z_N – zysk netto;

$$W_{RN} = \frac{Z_N}{K} \tag{11}$$

gdzie:
W_{RN} – wskaźnik rentowności netto,
Z_N – zysk netto,
K – koszt własny sprzedaży;

$$W_{PK} = \frac{K}{P} \tag{12}$$

gdzie:
W_{PK} – wskaźnik poziomu kosztów,
K – koszt własny sprzedaży,
P – przychody ze sprzedaży;

f) badanie źródeł pochodzenia i kierunków wykorzystywania środków pieniężnych.

Analiza finansowa dostarcza wiele użytecznych informacji odnośnie działalności przedsiębiorstwa oraz jego efektywności ekonomicznej, jednak występuje wiele ograniczeń, które powodują, że są to informacje nieścisłe. Niektóre przyczyny tych ograniczeń są następujące[12]:

- inflacja, która zakłóca zestawienie bilansowe przedsiębiorstwa (rozbieżności między wartościami zapisanymi a stanem faktycznym),
- sezonowość produkcji; dotyczy to pewnych sektorów działalności przedsiębiorstwa i powoduje błędy w analizie grupy współczynników obrotowych,
- trudności z oceną wartości miernika; czy jest on dobry czy zły, przykładem mogą być wartości współczynników płynności finansowej,

[12]Por. Brigham E.F., *Zarządzanie finansami*, PWE, Warszawa 1996, s. 87–88.

- trudności z oceną całej kondycji przedsiębiorstwa, wskutek występowania obok siebie sprzecznych wartości współczynników.

Z powyższych rozważań wynika, że ocena efektywności ekonomicznej oparta na analizie współczynnikowej nie jest pozbawiona słabych stron.

Ekonometryczne ujęcie problemu gospodarności

Miary zaprezentowane w poprzednim rozdziale nie pozwalają mierzyć efektów netto zarządzania, czyli nie pozwalają na oddzielenie w wynikach ekonomicznych dobrej pracy decydentów od działania czynników obiektywnych. Efekt netto zarządzania jest tym efektem gospodarowania, który można przypisać zdolnościom kierowniczym osób zarządzających, a nie warunkom, które zostały utworzone przez poprzedników bądź warunki naturalne[13]. W zależności od tego, jak decydenci wypełniają swoje zadania, osiągana jest skuteczność realizacji celów organizacji. Od decydenta wymaga się, aby jego działalność była jednocześnie sprawna, co oznacza, że decydent musi posiadać umiejętność właściwego działania, oraz skuteczna, czyli odznaczająca się umiejętnym wyborem właściwych celów.

Dysponując odpowiednio dopasowanym modelem ekonometrycznym zachowania się j-tych obiektów sterowania, istnieje możliwość wyznaczenia dla każdego z tych obiektów oceny wykorzystania zaangażowanych, istotnie oddziałujących czynników i warunków gospodarowania. Zależność tę można zapisać w następujący sposób:

$$y_j - \hat{y}_j = \hat{u}_j \tag{13}$$

gdzie:

$\hat{u}_j$ – miara efektywności rozpatrywanego procesu dla j-tego obiektu,

y_j – wykazywana, rzeczywista wartość funkcji celu (poziomu gospodarowania) uzyskana przez dany obiekt,

$\hat{y}_j$ – wartość teoretyczna (nadzieja matematyczna) poziomu gospodarowania, którą j-ty obiekt mógłby osiągnąć, prowadząc w sposób zgodny z oszacowaną funkcją regresji działalność produkcyjną.

---

[13]Por. Budziński R., *Identyfikacja i analiza postępu w efektywności gospodarowania*, [w:] System Naczelnego Kierownictwa w zarządzaniu (studia, algorytmy, modele), Wyd. Informa, Szczecin 1997, s. 107–108.

W ujęciu ekonometrycznym miarą sprawności gospodarowania jest podstawowa zależność metody MNK, a mianowicie: powrotne odniesienie uzyskanych wyników w j-tych obiektach do odpowiednio dopasowanego modelu zachowania się całej zbiorowości (funkcji regresji), gdzie wariancja resztowa s^2 osiąga minimum. Należy przy tym wyjaśnić, kiedy różnice dla dowolnej i-tej funkcji celu są istotne, tzn. kiedy oszacowana wartość y_{ji} istotnie różni się od wartości zrealizowanej y_{ji}. Posłużono się statystyką podaną przez Budzińskiego i Kopcia [1981]:

$$\frac{y_j - \hat{y}_j}{S_e} \quad \text{dla} \quad j = 1, 2, ..., n, \tag{14}$$

gdzie S_e jest szacunkiem wartości odchylenia standardowego składnika losowego. Za pomocą tej statystyki możliwa jest weryfikacja założeń hipotezy, H_0: $y_j = \hat{y}_j$ wobec H_1: $y_j \neq \hat{y}_j$. Licznik jest zmienną o rozkładzie normalnym, natomiast mianownik jest pierwiastkiem kwadratowym ze zmiennej o rozkładzie x^2_{n-k-1} dla poziomu istotności α i $n-k-1$ stopni swobody. Statystyka $(y_j - \hat{y}_j)/S_e$ ma rozkład *t-Studenta*. Oznaczając przez t_α wartości krytyczne odczytywane z tablic rozkładu *t-Studenta* dla poziomu istotności α (zakłada się z reguły $\alpha = 0{,}05$) i $n-k-1$ stopni swobody, utworzone zostają następujące klasy sprawności (kolejno skuteczności oraz efektywności), mianowicie: niezwykle niska (NN), niska (N), przeciętna (P), wysoka (W) i wyjątkowo wysoka (WW), gdy:

(NN) $y_j - \hat{y}_j \leq -S_e t_\alpha$, (15)

(N) $-S_e t_\alpha < y_j - \hat{y}_j \leq -S_e$, (16)

(P) $-S_e < y_j - \hat{y}_j \leq S_e$, (17)

(W) $S_e < y_j - \hat{y}_j \leq S_e t_\alpha$, (18)

(WW) $y_j - \hat{y}_j > S_e t_\alpha$. (19)

Przedstawione powyżej ujęcie klasyfikacji sprawności wynika z założenia, że elementem głównym kreowania dyspersji wyników jest oszacowane resztowe odchylenie standardowe (S_e). Mówi ono, o ile średnio różnią się rzeczywiste wyniki (np. przychody przedsiębiorstw y_i) od wyników wzorcowych $\hat{y}_j$, które wynikają z oszacowanej postaci równania regresji. Warto przytoczyć, że przy

założeniu normalnego rozkładu reszt $y_j - \hat{y}_j$ około 66% całej ich populacji mieści się w przedziale $(-S_e, S_e)$, a blisko 96% w przedziale $(-2S_e, 2S_e)$, co w przybliżeniu może tworzyć prawdopodobieństwo wystąpienia: NN ≈ 2%, N ≈ 15%, P ≈ 66%, W ≈ 15% i WW ≈ 2%.

Podobnie jak sprawność można ocenić także skuteczność w osiąganiu maksymalnego stopnia realizacji celu. Istotnym momentem tej klasyfikacji jest wartość odchylenia standardowego (S_y) zmiennej zależnej y_i, stanowiąca kryterium oceny oraz średnia ($\bar{y}$) poziomu y_i dla danej grupy jednostek gospodarujących w badanej próbie. Przyjęta wartość krytyczna t^*_α jest odczytana z tablic rozkładu *t-Studenta* dla $n - 1$ stopni swobody i poziomu istotności α.

(NN) $y_j - \bar{y}_j \leq -S_y t^*_\alpha$, (20)

(N) $-S_y t^*_\alpha < y_j - \bar{y}_j \leq -S_y$, (21)

(P) $-S_y < y_j - \bar{y}_j \leq S_y$, (22)

(W) $S_y < y_j - \bar{y}_j \leq S_y t^*_\alpha$, (23)

(WW) $y_j - \bar{y}_j > S_y t^*_\alpha$. (24)

Rozstrzygnięcie o efektywności danego obiektu wymaga oceny skuteczności i sprawności. Ważne przy tym jest ujęcie zmiennych preferencji analityka-decydenta w ocenie miejsca tych kryteriów (rangi – ważności) w całościowym spojrzeniu na analizowane obiekty i otaczającą je rzeczywistość. Pewnym rozwiązaniem może tu być połączenie przyjętych przedziałów klasowych, czyli sprawności i skuteczności, aby dokonywać wspólnej oceny jednocześnie według obu kryteriów. Jedną z możliwych koncepcji takiego połączenia jest proste zsumowanie lewych i prawych stron odpowiednich nierówności. Dla uwzględnienia możliwości przypisania różnej skali wpływu obu kryteriów na efektywność wprowadzono nadawanie wag (rang) dla częstego dylematu: sprawny – skuteczny.

Należy przyjąć, że $y_j - \hat{y}_j = S_p$, a $y_j - \bar{y}_j = S_k$ oraz K to ranga dla skuteczności i P ranga dla sprawności. Używając wyżej przedstawionych oznaczeń, ważone kryterium oceny efektywności można zapisać w następującej formie:

(NN) $KS_k + PS_p \leq -(PS_e t_\alpha + KS_y t^*_\alpha)$, (25)

(N) $-(PS_e t_\alpha + KS_y t^*_\alpha) < KS_k + PS_p \leq -(PS_e + KS_y)$, (26)

$$(P) - (PS_e + KS_y) < KS_k + PS_p \leq PS_e + KS_y, \quad (27)$$

$$(W)\ PS_e + KS_y < KS_k + PS_p \leq PS_e t_\alpha + KS_y t^*_\alpha, \quad (28)$$

$$(WW)\ KS_k + PS_p > PS_e t_\alpha + KS_y t^*_\alpha. \quad (29)$$

Należy zwrócić uwagę, że wartości K i P są liczbami naturalnymi, a istotne znaczenie ma jedynie ich wzajemny stosunek, który może odzwierciedlać preferencje analityka (decydenta) w ocenie funkcjonowania, np. grupy przedsiębiorstw. Przyjęcie odpowiedniej wartości rang K i P jest uzależnione od tego, jaki pogląd reprezentuje użytkownik realizujący badania efektywności. Łatwo przy tym zauważyć, że jeśli założy się $K = 0$ i $P = 1$, to jest to badanie sprawności firmy (przedsiębiorstwa), natomiast dla $K = 1$ i $P = 0$ badanie skuteczności, a dla innych przypadków są to różne warianty pośrednie.

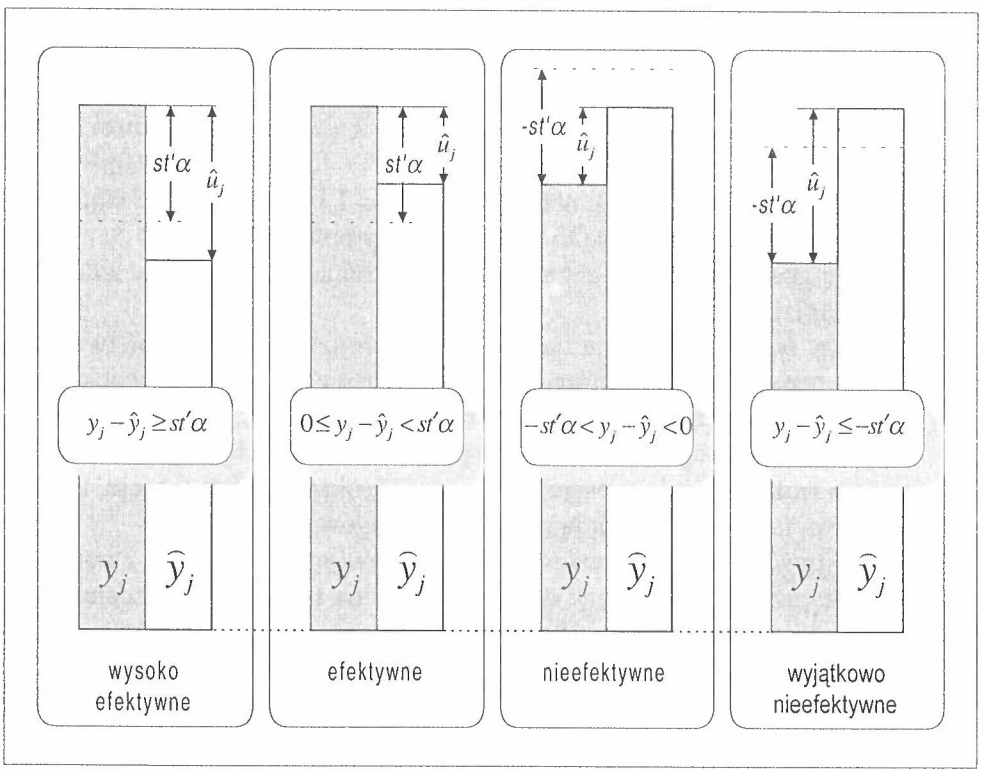
Analizę gospodarności można prowadzić trzema sposobami, w zależności od celu badania:

- pierwszy sposób polega na badaniu gospodarności przedsiębiorstwa na tle innych przedsiębiorstw, które należą do tej samej gałęzi; informacje zastosowane w tym przypadku mają charakter przekrojowy,
- drugi sposób polega na analizie gospodarności przedsiębiorstwa w danym okresie (rok, kwartał, miesiąc) na tle innych okresów; informacje, które są podstawą takiej analizy mają charakter szeregów czasowych,
- trzeci sposób jest połączeniem dwóch omówionych powyżej i polega na analizie przedsiębiorstwa w danym okresie na tle innych przedsiębiorstw i na tle innych okresów; w tym przypadku ma się do czynienia z danymi przekrojowo-czasowymi[14].

Rozpatrywanie problemu efektywności gospodarowania w czasie związane jest z oszacowaniem tendencji rozwojowych i stabilnością w postępie efektywności gospodarowania. Przede wszystkim chodzi o zastanowienie się nad stabilnością procesu gospodarowania na tle ukształtowanej tendencji rozwojowej. Jeśli w którymś z obiektów oszacowano poziom gospodarowania w ujęciu statycznym, to należy się zastanowić, czy jest to przejaw normalnej tendencji rozwojowej, czy też chwilowy (wybuchowy) postęp lub regres w prowadzeniu przedsiębiorstwa. Chodzi tu o zjawisko homeostazy, czyli dynamicznego przywracania celowo naruszanej równowagi wewnętrznej systemu. Jest to znaczący moment dla postępu w efektywności gospodarowania. W postępie tym ważna jest zdolność tego procesu (poprzez trafne decyzje) do przywracania naruszanej

[14]Por. Hozer J., *Zastosowanie metod ekonometrycznych w analizie gospodarności przedsiębiorstw*, Prace Naukowe PS nr 75, Szczecin 1978, s. 15.

dynamicznie równowagi wewnętrznej w wymianie energii i informacji z otoczeniem przedsiębiorstwa. Cecha ta pozwala odróżnić układy inteligentne, które mają zdolność identyfikowania i kojarzenia interakcji pomiędzy stanami elementów nastawialnych a wynikami gospodarowania, od układów nie tworzących postępu (regulujących), występujących najczęściej w automatyce[15].



Rysunek 1
Graficzna postać wyznaczania efektywności
Źródło: Budziński R., op. cit., s. 115.

[15]Por. Budziński R., *Komputerowy system przetwarzania danych ekonomiczno-finansowych w przedsiębiorstwie*, Wyd. Naukowe Uniwersytetu Szczecińskiego, Warszawa-Szczecin 2000, s. 236.

Zastosowanie metod sztucznej inteligencji

W praktyce decydowania występują zależności, których poprawne oszacowanie wymaga często innych metod niż te, gdzie identyfikacji podlegają mierzalne (ilościowe) związki przyczynowo-skutkowe. Mogą one mieć charakter jakościowy (niemierzalny). Znajomość uwarunkowań w postępie efektywności gospodarowania jest jedną z zasadniczych przesłanek podejmowania trafnych decyzji gospodarczych. Duże możliwości może dać (jak się wydaje) zastosowanie metod sztucznej inteligencji (sztucznych sieci neuronowych) do identyfikacji i obiektywizacji głównego dylematu przedstawionego artykułu: ile jest, a ile być powinno?

Sieć neuronowa jest bardzo uproszczonym modelem mózgu. Składa się z dużej liczby (od kilkuset do kilkudziesięciu tysięcy) elementów, które przetwarzają informację. Elementami tymi są neurony powiązane w sieć za pomocą połączeń o parametrach (zwanych wagami) modyfikowanych w trakcie tak zwanego procesu uczenia. Topologia połączeń oraz ich parametry stanowią program działania sieci. Natomiast sygnały, które pojawiają się na jej wyjściach, będące odpowiedzią na określone sygnały wejściowe, są rozwiązaniami stawianych jej zadań[16]. Sieci neuronowe uważane są za naturalne aproksymatory funkcji. Do modelowania najczęściej stosuje się jednokierunkowe wielowarstwowe sieci, zwane perceptronami wielowarstwowymi, z sigmoidalnymi funkcjami aktywności neuronów. Znając wszystkie wejścia objaśniające systemu, nauczanie sieci polega na znalezieniu jej optymalnej struktury (liczba warstw, liczba neuronów w warstwie) oraz optymalnej wartości parametrów w neuronach. Podstawą działania sieci są algorytmy uczące, umożliwiające zaprojektowanie odpowiedniej struktury sieci i dobór parametrów tej struktury, dopasowanych do problemu podlegającego rozwiązaniu.

Jednym z kryteriów podziału sieci neuronowych jest podział na sieci z jednokierunkowymi połączeniami (*feedforward*) i ze sprzężeniami zwrotnymi (*sieci Hopfielda*). Cechą charakterystyczną sieci jest zdolność do adaptacji i samorealizacji oraz mała wrażliwość na uszkodzenia elementów, a także zdolność do równoległej pracy.

Sieć neuronowa uczy się na dwa sposoby. Pierwszy z nich i jednocześnie najczęściej stosowany to *uczenie pod nadzorem* (uczenie z nauczycielem, np. metoda propagacji wstecznej). Próbka zbioru uczącego określa wszystkie wej-

---

[16]Por. Tadeusiewicz R., *Sieci neuronowe*, Akademicka Oficyna Wydawnicza, Warszawa 1993, s. 13.

ścia oraz wyjścia, które są wymagane przy prezentacji tych danych wejściowych. Kolejnym krokiem jest wybór podzbioru zbioru uczącego i podanie próbki z tego podzbioru na wejście sieci. Dla każdej próbki jest porównywalny aktualny sygnał wejściowy sieci z wyjściowym. Po przetworzeniu całego podzbioru próbek uczących następuje korygowanie wag łączących neurony w sieć. Drugim sposobem uczenia jest *uczenie bez nadzoru* (uczenie bez nauczyciela, np. algorytm uczenia konkurencyjnego). W tym przypadku istnieje także zbiór próbek sygnałów wejściowych. Różnica polega na tym, że sieć nie jest doprowadzana do pożądanych odpowiedzi wyjściowych na te próbki. Proces uczenia polega na umożliwieniu wykrycia istotnych cech zbioru uczącego i wykorzystanie ich do grupowania sygnałów wejściowych na klasy, które sieć potrafi rozróżniać. Istnieje także trzeci sposób uczenia sieci, czyli *uczenie ze wzmocnieniem*. Metoda ta zawiera w sobie elementy dwóch opisanych powyżej metod. Jest jednocześnie uczeniem bez nadzoru, ponieważ pożądany sygnał wyjściowy nie jest określony, oraz uczeniem pod nadzorem, gdyż jeśli sieć odpowiada na próbkę ze zbioru treningowego, to wiadomo, czy odpowiedź ta jest dobra czy nie.

Sztuczne sieci neuronowe mają wiele właściwości pożądanych w zastosowaniach praktycznych, a mianowicie:

- stanowią uniwersalny układ aproksymacyjny, który odwzorowuje wielowymiarowe zbiory danych,
- mają zdolności uczenia i adaptacji do zmieniających się warunków środowiskowych,
- mają zdolność uogólniania nabytej wiedzy[17].

Mogą okazać się lepszymi od innych metod w sytuacjach, gdy:

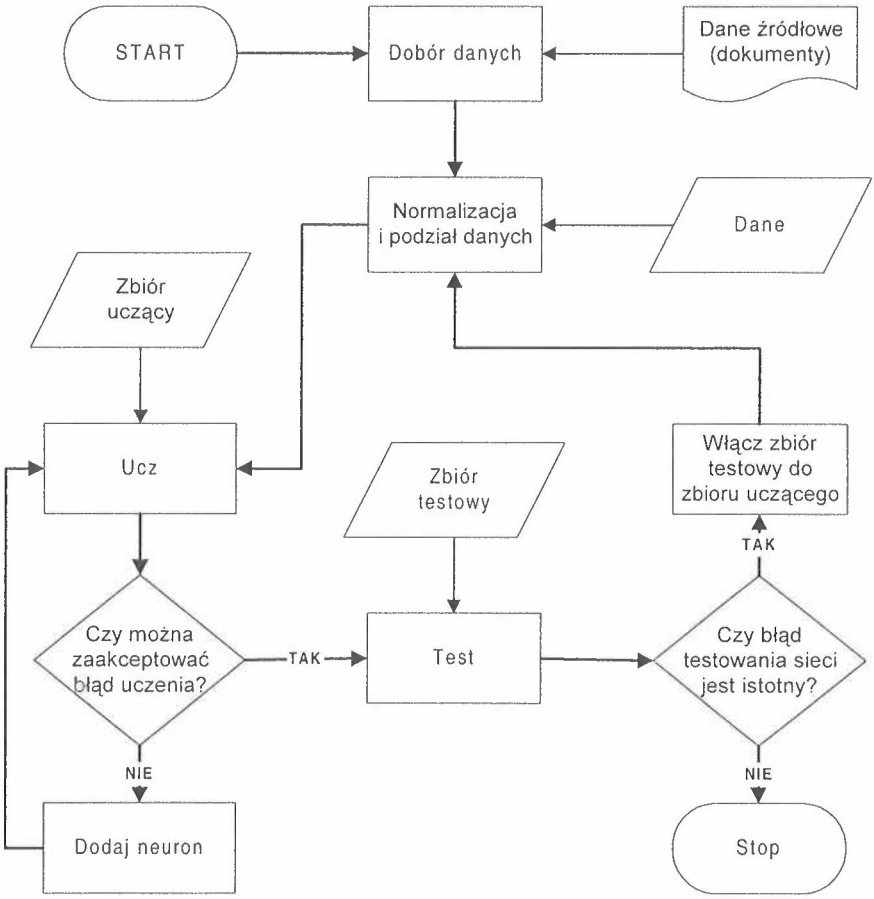
- dane, z których należy wyciągnąć wnioski, są „rozmyte", czyli jeśli danymi wejściowymi są ludzkie opinie bądź błędnie określone kategorie , bądź też dane obarczone są dość dużymi błędami,
- ważne do podjęcia wymaganej decyzji wzorce są głęboko ukryte, czyli niewykrywalne przez zmysły naukowców i tradycyjne metody statystyczne,
- dane wykazują znaczną, nieoczekiwaną nieliniowość,
- dane są chaotyczne (w sensie matematycznym)[18].

Dobre własności uogólniające sieci neuronowych sprawiają, że można je zastosować do rozwiązywania różnego rodzaju zadań.

---

[17]Por. Osowski S., *Sieci neuronowe w ujęciu algorytmicznym*, Wydawnictwa Naukowo-Techniczne, Warszawa 1996, s. 11.

[18]Por. Masters T., *Sieci neuronowe w praktyce. Programowanie w języku C++*, Wydawnictwa Naukowo-Techniczne, Warszawa 1996, s. 21.



Schemat 3
Proces uczenia nadzorowanego w sieci
Źródło: Opracowanie własne.

Sieć neuronowa ma umiejętność selekcji ważnych parametrów i ignorowania nieznaczących parametrów. Dlatego gdy nie ma pewności, czy dany czynnik ma wpływ na problem, należy decyzję o jego znaczeniu pozostawić sieci neuronowej i włączyć go do zbioru treningowego. Podstawowym krokiem w procesie uczenia sieci jest umiejętne postawienie problemu.

Jednak, jak do tej pory, nie ma przykładów praktycznych zastosowań sztucznych sieci neuronowych w badaniach efektywności gospodarowania. Rozważania koncentrują się na przede wszystkim na podejściach opartych na różnego rodzaju klasyfikacji z udziałem teorii zbiorów rozmytych. Logikę roz-

mytą, będącą jednym z działów teorii zbiorów rozmytych, stosuje się do opisywania i przetwarzania informacji nieprecyzyjnych, których definicja wymaga nieścisłego podejścia. Model rozmyty składa się z trzech bloków:

- fuzyfikacji (rozmywanie); model rozmyty na wejściu ma wartości ostre (dokładne), modelowanie rozpoczyna się od przetworzenia tych wartości do postaci rozmytej; polega to na obliczeniu stopnia przynależności danego wejścia do zbiorów rozmytych, używając zdefiniowanych funkcji przynależności ($\mu(x)$); w wyniku tych działań otrzymuje się wartości, które informują o stopniu przynależności wejścia do poszczególnych zbiorów rozmytych,
- inferencji (wnioskowanie); w bloku tym następuje obliczenie wynikowej funkcji przynależności,
- defuzyfikacji (ostrzenie); na podstawie wynikowej funkcji przynależności obliczona jest ostra wartość, będąca rezultatem podania ostrych wartości na wejściu modelu.

Dane wyjściowe każdego z bloków są jednocześnie wejściem dla bloku następnego. Istnieje wiele metod modelowania i sterowania rozmytego, dlatego też trudno wybrać odpowiednie podejście do badanego zagadnienia.

Według Plucińskiego i Piegata [1999], kompleksowa ocena przedsiębiorstw jest typowym zagadnieniem rozmytym – operuje „ziarnami informacji" typu „bardzo dobra", „średnia", „zła"[19]. Autorzy cytowanej w artykule pracy zaprezentowali podejście, w którym analizę dyskryminacyjną, polegającą na ocenie przedsiębiorstwa opartej na wygospodarowanych nadwyżkach pieniężnych, zastąpili metodami opartymi na teorii zbiorów rozmytych. Zastosowali metodę klasyfikacji rozmytej (wersję opracowaną przez A. Piegata), opartą na projekcji wielowymiarowej funkcji rozmytej opisującej przynależność próbki pomiarowej do danej klasy na podprzestrzenie dwuwymiarowe (w skrócie: klasyfikacja metodą projekcji).

W metodzie tej przyjmuje się, że n-wymiarowa funkcja przynależności jest nieregularną „górą" o maksymalnej wysokości równej 1. Jeżeli próbka pomiarowa należy do danej klasy, to jej stopień przynależności zarówno do n-wymiarowej funkcji przynależności, jak i do jej 2-wymiarowych projekcji musi być większy od 0. Oznacza to, że 2-wymiarowe rzuty próbek pomiarowych muszą leżeć w obrębie 2-wymiarowego rzutu n-wymiarowej funkcji przynależności do danej klasy. Dzięki temu można oddzielnie w różnych podprzestrzeniach konstruować funkcje przynależności, a następnie na ich podsta-

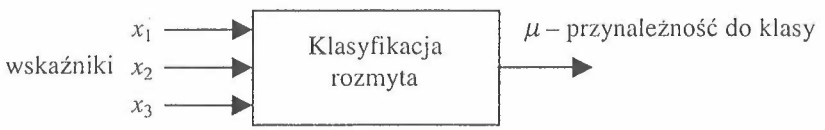
---

[19]Por. Pluciński M., Piegat A., *Rozmyta klasyfikacja przedsiębiorstw metodą projekcji przestrzeni wielowymiarowej na podprzestrzenie 2-wymiarowe*, [w:] Informatyka i zarządzanie strategiczne (red. – Budziński R.), Wyd. Informa, Szczecin 1999, s. 448.

wie skonstruować pełną n-wymiarową funkcję przynależności według poniżej podanego wzoru[20].

$$\mu(A_{1 \div n}) = \mu(A_{12} \wedge A_{34} \wedge ... \wedge A_{n-1,n}) \qquad (30)$$

Przed rozpoczęciem konstruowania funkcji przynależności do klas należy dokonać normalizacji wskaźników do zakresu [0, 1].



Rysunek 2
Klasyfikacja rozmyta w formie „black box"
Źródło: Pluciński M., Piegat A., op. cit., s. 450.

Klasyfikacja rozmyta metodą projekcji umożliwia rozłożenie problemu wielowymiarowego na 2-wymiarowe. Dzięki temu w 2-wymiarowych podprzestrzeniach można stosować bardziej złożone funkcje przynależności o większej liczbie stopni swobody, które lepiej dostosowują się do rozkładu próbek pomiarowych poszczególnych klas. Strojenie wolnych parametrów może być realizowane wizualnie przez obserwację położenia funkcji przynależności i położenia próbek na ekranie monitora.

W modelowaniu rozmytym można napotkać pewne ograniczenia i trudności. Są nimi między innymi liczba i jakość wejść systemu. Ma się na myśli to, że niektóre wejścia są bardziej, inne natomiast mniej istotne dla modelowanego systemu, co prowadzi do wzrostu bądź spadku jakości generowanego wyjścia. Dlatego już na wstępie dobrze jest zastanowić się nad właściwymi wejściami, eliminując te, które są nieistotne dla modelu. Innym problemem jest zbyt duża liczba wejść, którą należy zminimalizować metodą prób i błędów bądź metodą krzywych rozmytych.

[20]Por. Pluciński M., Piegat A., op. cit., s. 448.

Uwagi końcowe

Podsumowując rozważania na temat efektywności gospodarowania, należy stwierdzić, że jest ona jedną z głównych kategorii ekonomicznych, objaśnianą za pomocą relacji między różnymi postaciami uzyskiwanych wyników zarówno produkcyjnych, jak i finansowych oraz najlepiej opisującą rezultaty racjonalnego gospodarowania w przedsiębiorstwie. W praktyce do oceny efektywności gospodarowania stosuje się najczęściej kryteria w postaci różnicowej, czyli zysku, oraz ilorazowej (rentowność), a także wykorzystuje się specjalne zespoły indeksów efektywności. Miary te odgrywają znaczącą rolę w zarządzaniu, gdzie są podstawą wyboru decyzji. Są także regulatorami pobudzania i zasilania systemu motywacyjnego pracy. Nie wyjaśniają jednak postawionego problemu badawczego, tj. analizy gospodarności w konkretnych warunkach gospodarowania dla konkretnego przedsiębiorstwa. Istniejące w gospodarce rynkowej systemy oceny efektywności ekonomicznej przedsiębiorstwa, oparte na analizie współczynnikowej, nie pozwalają na przeprowadzenie obiektywnej oceny funkcjonowania przedsiębiorstwa. Ocena efektywności ekonomicznej przedsiębiorstwa powinna być wieloaspektowa, uwzględniająca łączenie wielu elementów ją determinujących. Natomiast przeprowadzenie analizy tylko na podstawie metod tradycyjnych, za pomocą jednego miernika powoduje słabą wartość poznawczą, a z kolei większa liczba wskaźników daje rozbieżności co do całościowej oceny funkcjonowania przedsiębiorstwa. Właściwe rozwiązania możemy uzyskać za pomocą analizy ekonometrycznej. Metoda ta nie jest jednak bez wad. Niezmiernie trudno jest znaleźć postać równania, które równocześnie spełnia kryteria merytoryczne doboru zmiennych z równoczesną formalną estymacją zgodności (np. z rozkładem normalnym) i eliminacją autokorelacji. Ważnym momentem jest istotność korelacji i parametrów regresji. Ponadto, w praktyce decydowania występują zależności, których poprawne oszacowanie wymaga często innych metod lub mają one charakter jakościowy (niemierzalny). Znajomość uwarunkowań w analizie efektywności gospodarowania jest jedną z zasadniczych przesłanek podejmowania trafnych decyzji gospodarczych. Duże nadzieje wiąże się z możliwościami (jak się wydaje) zastosowania metod sztucznej inteligencji do identyfikacji i obiektywizacji głównego dylematu przedstawionego artykułu: ile jest, a ile być powinno. Problem jednak jest w tym, że w metodach sztucznej inteligencji nie można – jak w metodach analizy ekonometrycznej – oszacować postaci równania funkcji regresji, tj. praw rządzących działaniem czynników, które w istotny sposób wpływają (i pozwalają oszacować) na to, kto jest gospodarny (i dlaczego), a kto nim nie jest.

Literatura

1. BRIGHAM E.F., *Zarządzanie finansami*, PWE, Warszawa 1996.
2. BOCHNAK M., *Opracowanie oprogramowania do klasyfikacji przedsiębiorstw z wykorzystaniem logiki rozmytej*, Praca dyplomowa, Szczecin 2000.
3. BUDZIŃSKI R., KOPEĆ J., *Econometric Model of Activity as a Diagnostic Tool Analysing Agriculrural Enterprices of a Region*, [w:] Spatial Development: Elements of Systems Analytic Approach, IBS PAN, Warszawa 1981.
4. BUDZIŃSKI R., *Identyfikacja i analiza postępu w efektywności gospodarowania*, [w:] System Naczelnego Kierownictwa w zarządzaniu (studia, algorytmy, modele), Wyd. Informa, Szczecin 1997.
5. BUDZIŃSKI R., *Komputerowy system przetwarzania danych ekonomiczno-finansowych w przedsiębiorstwie*, Wyd. Naukowe Uniwersytetu Szczecińskiego, Warszawa-Szczecin 2000.
6. CZECHOWSKI L., *Wielowymiarowa ocena efektywności ekonomicznej przedsiębiorstwa przemysłowego*, Wyd. Uniwersytetu Gdańskiego, Gdańsk 1997.
7. GRZESIAK S., *Metody ilościowe w badaniu efektywności ekonomicznej przedsiębiorstw*, Wyd. Naukowe Uniwersytetu Szczecińskiego, Rozprawy i Studia T. (CCCXL) 226, Szczecin 1997.
8. HOZER J., *Zastosowanie metod ekonometrycznych w analizie gospodarności przedsiębiorstw*, Prace Naukowe PS nr 75, Szczecin 1978.
9. HOZER J., ZAWADZKI J., *Proces ekonometrycznego modelowania*, US, Rozprawy i Studia T./CXXV/51, Szczecin 1990.
10. KORBICZ J., OUCHOWICZ A., UCIŃSKI D., *Sztuczne sieci neuronowe. Podstawy i zastosowania*, Akademicka Oficyna Wydawnicza PLJ, Warszawa 1994.
11. MASTERS T., *Sieci neuronowe w praktyce. Programowanie w języku C++*, Wydawnictwa Naukowo-Techniczne, Warszawa 1996.
12. OSOWSKI S., *Sieci neuronowe w ujęciu algorytmicznym*, Wydawnictwa Naukowo-Techniczne, Warszawa 1996.
13. PLUCIŃSKI M., PIEGAT A., *Rozmyta klasyfikacja przedsiębiorstw metodą projekcji przestrzeni wielowymiarowej na podprzestrzenie 2-wymiarowe*, [w:] Informatyka i zarządzanie strategiczne (red. – Budziński R.), Wyd. Informa, Szczecin 1999.
14. *Słownik ekonomiczny*, Wyd. ZNCZ, Szczecin 1994.
15. STACHAK S., *Ekonomika agrofirmy*, PWN, Warszawa 1998.

16. TADEUSIEWICZ R., *Sieci neuronowe*, Akademicka Oficyna Wydawnicza, Warszawa 1993.
17. URBAŃCZYK E., *Metody analizy ekonomicznej efektywności majątku trwałego w przemyśle*, Prace Naukowe Politechniki Szczecińskiej nr 300, Szczecin 1985.
18. WAŚNIEWSKI T., SKOCZYLAS W., *Zasady analizy finansowej w praktyce. Przykłady i zadania*, Fundacja Rozwoju Rachunkowości w Polsce, Warszawa 1994.

Measurement methods of management effectiveness in an organisation

Abstract

In the article the three methodical approaches to the issue of measurement of management effectiveness are presented. They are as follow: measure of economic analysis, measure of econometric analysis, artificial intelligence methods. Stress has been put mainly on the effect of management which has been achieved by managerial abilities of managers rather than conditions created by their predecessors or natural conditions. The effectiveness of organisation goals achievement depends on how the decision makers undertake their tasks, another words it depends on their abilities (managerial talent, energetic and reasonable work).

Danuta Bogocz
Zakład Statystyki Matematycznej AR w Krakowie

Analiza zastosowań teorii zbiorów rozmytych w badaniach ekonomicznych na wybranych przykładach

Wstęp

Pojęcia matematyczne, z jednej strony będąc w pełni autonomicznymi tworami abstrakcyjnymi, stanowią jednocześnie podstawę do rozwoju teorii znajdujących zastosowanie w różnych dziedzinach, niekoniecznie bezpośrednio związanych z ich wyjściowym pochodzeniem. Kontekst ten niewątpliwie wpływa na kształt rozwijanej teorii, koncentrując uwagę na tych aspektach bądź cechach danego pojęcia, które są tu szczególnie przydatne. Rezultatem tego rodzaju analiz są często nowe teorie, powstałe w wyniku modyfikacji definicji, odrzucenia aksjomatu itp.

Geneza pojęcia zbioru rozmytego sięga 1965 roku. Punktem wyjścia do rozwoju teorii jest definicja zbioru za pomocą funkcji charakterystycznej.

Definicja 1. Funkcja $\mu A : X \rightarrow [0, 1]$ jest funkcją charakterystyczną zbioru A wtedy i tylko wtedy, gdy dla wszystkich x

$$\mu_A(x) = \begin{cases} 1, & \text{dla } x \in A, \\ 0, & \text{dla } x \notin A. \end{cases}$$

Teoria zbiorów rozmytych stanowi rozszerzenie tak ujętej teorii zbiorów. Zdefiniowany wyżej zbiór jest tam nazywany zbiorem ostrym, gdyż dla dowolnego elementu przestrzeni X mamy: $x \in A$ albo $x \notin A$. Obok teorii zbiorów rozmytych, będącej najbardziej upowszechnionym uogólnieniem własności przynależności (tj. funkcji charakterystycznej), znane są inne próby jej modyfikacji, takie jak wstępna teoria zbiorów czy teoria Dempstera-Shafera.

Definicja zbioru rozmytego za pomocą funkcji przynależności jest następująca:

Definicja 2.1. Funkcja przynależności μ_A zbioru rozmytego A określonego w przestrzeni U jest funkcją postaci

$$\mu_F : U \rightarrow [0, 1].$$

Gwoli ścisłości trzeba tu dodać, że w ujęciu czysto formalnym teoria ta jest właściwie teorią podzbiorów rozmytych. Niemniej biorąc pod uwagę fakt, że każdy podzbiór z definicji sam w sobie jest zbiorem, możemy dla uproszczenia stosować te określenia zamiennie. Dalej podajemy szczegółową definicję podzbioru rozmytego przytaczaną przez C. Noworola za jednym z głównych prekursorów pojęcia, L.A. Zadehem (zob. [8]).

Definicja 2.2. Niech U będzie zbiorem skończonym lub nieskończonym. Podzbiór rozmyty A zbioru U określa formuła:

$$A=\{(x: \mu_A(x)\in \mathrm{M}) \text{ , dla każdego } x\in U\},$$

gdzie:
$\mu_A(x)$ oznacza stopień przynależności elementu x do zbioru rozmytego A,
μ_A oznacza funkcję przynależności, która każdemu elementowi x przyporządkowuje pewną wartość ze zbioru M.
Zbiór M nazywa się zbiorem przynależności. Zakłada się, że jest to zbiór uporządkowany. W większości przypadków przyjmuje się przedział [0, 1] (zob. definicja 2.1.).

W niektórych pozycjach literatury, zwłaszcza z zakresu zastosowań teorii zbiorów rozmytych w badaniach operacyjnych i praktyce gospodarczej, można napotkać nieco inną definicję zbioru rozmytego jako liczby rozmytej.

Definicja 2.3. Liczbą rozmytą nazywamy uporządkowaną czwórkę liczb rzeczywistych dowolnego znaku $\left(\underline{a},\underline{b},\overline{b},\overline{a}\right)$, gdzie $\underline{a} \le \underline{b} \le \overline{b} \le \overline{a}$.

Definicja 2.4. Liczbę rozmytą $\left(\underline{a},\overline{a},\alpha,\beta\right)_{L\text{-}R}$ nazywamy liczbą typu L-R, jeśli jej funkcja uczestnictwa ma postać:

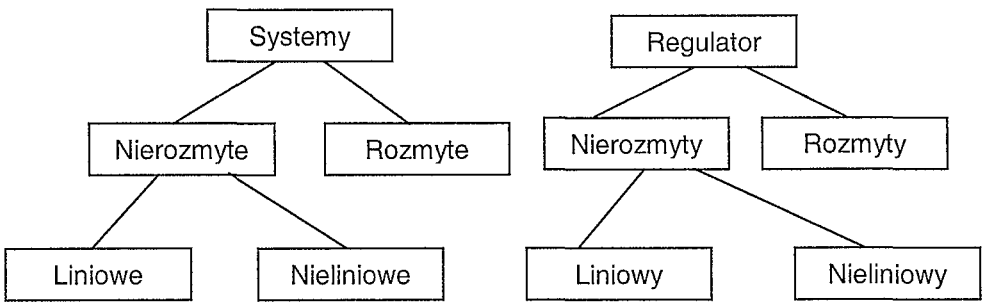
$$\mu_A(x)=\begin{cases} L((\underline{a}-x)/\alpha), & \text{dla } x\in [-\infty,\underline{a}] \\ 1, & \text{dla } x\in [\underline{a},\overline{a}], \\ R((x-a)/\beta) & \text{dla } x\in [\overline{a},\infty] \end{cases}$$

gdzie L i R są funkcjami kształtu, zaś α i β – nieujemnymi parametrami.

Definicja 2.3 stanowi pewne zawężenie rodziny liczb rozmytych do tzw. *trapezowych liczb rozmytych*, niemniej na potrzeby niniejszej pracy zakres ten jest w zupełności wystarczający.

Zbiory rozmyte w teorii sterowania

Pojęcie zbioru rozmytego najwcześniej zostało zaadaptowane na gruncie teorii systemów. Jedną z gałęzi tej szerokiej i prężnie rozwijającej się dziedziny jest teoria sterowania. Tu również w ostatnich czasach coraz większą popularność zyskuje tzw. sterowanie rozmyte. Przyczyn tego zjawiska jest wiele. Jedną z nich jest nieliniowość obiektów sterowania, gdyż procesy rzeczywiste wymagające automatycznego sterowania są z reguły procesami nieliniowymi. W zależności od rodzaju nieliniowości stosowane są różne typy regulatorów rozmytych opartych na wiedzy, nazywanych w skrócie regulatorami FKBC (ang. *Fuzzy Knowledge Based Controllers*). Na rysunku 1 schematycznie przedstawiono systemy i regulatory.



Rysunek 1
Uogólniony schemat systemów i regulatorów

Podstawowe elementy tworzące strukturę regulatora FKBC to:

1. Moduł rozmytości. Tu parametrem projektowania jest wybór strategii fuzyfikacji.
2. Baza wiedzy, na którą składają się:
   - baza danych – nośnik niezbędnych informacji zawierających zbiory rozmyte odwzorowujące znaczenie wartości lingwistycznych stanu

procesu i zmiennych sterujących oraz fizyczne dziedziny i ich znormalizowane odpowiedniki,
- baza reguł – strukturalne przedstawienie strategii sterowania.

1. Maszyna wnioskująca.
2. Moduł defuzyfikacji. Tu parametrem projektowania jest wybór operatorów defuzyfikacji.

Wśród regulatorów FKBC na szczególną uwagę zasługują regulatory adaptacyjne. Regulatory te przestrajane są automatycznie w zależności od bieżących charakterystyk procesu, co wobec zmienności parametrów procesu wraz ze zmianą punktu pracy stanowi istotny atut w porównaniu z regulatorami konwencjonalnymi, każdorazowo dostrajanymi w danym punkcie pracy lub w ograniczonym przedziale czasu.

Elementami różniącymi regulatory adaptacyjne od regulatorów konwencjonalnych są monitor procesu oraz mechanizm adaptacyjny. Funkcję monitora może spełniać miara jakości szacująca jakość sterowania lub estymator parametru aktualizujący model procesu. Z kolei mechanizm adaptacyjny wykorzystuje informację dochodzącą z monitora procesu do aktualizacji parametrów regulatora, adaptując tym samym regulator do zmieniających się charakterystyk procesu. Parametrami, które mogą być zmieniane w celu modyfikacji działania regulatora są:
- czynniki skalujące dla każdej zmiennej,
- zbiór rozmyty odwzorowujący znaczenie wartości lingwistycznych,
- reguły JEŻELI-TO.

Regulatory adaptacyjny zastosowano m.in. do symulacji komputerowej zbiornika masy, pieca cementowego oraz wyparki z pojedynczą wymuszoną cyrkulacją.

Zbiory rozmyte w badaniach operacyjnych

Badania operacyjne w praktyce gospodarczej

Pewne ciekawe propozycje rozmytych rozwiązań problemów oceny i wyboru inwestycji można znaleźć w pracy Kuchty [6]. Autorka przedstawia rozszerzenia metod klasycznych na przypadek nieprecyzyjnych danych, w szczególności nieprecyzyjnej znajomości momentów, w których będą następowały przepływy gotówkowe oraz wzrostu niepewności w ocenie stopy oprocentowania z upływem czasu, koncentrując uwagę na metodzie teraźniejszej wartości netto (ang. *Net Present Value*), nazywanej w skrócie NPV. Wskaźnik wzrostu wartości firmy dzięki realizacji danego projektu inwestycyjnego w metodzie NPV jest postaci:

$$NPV = CF_0 + \sum_{j=1}^{n} \frac{CF_j}{(1+r)^j}, \qquad (*)$$

gdzie:
n – przewidywany czas trwania projektu,
CF_j (j = 0, ..., n) – przepływy gotówkowe (nakłady lub wpływy) związane z danym projektem inwestycyjnym, mające miejsce na końcu j-tego roku,
r – stopa dyskontowa, wyznaczana na podstawie kosztu kapitału potrzebnego do pozyskania środków na projekt inwestycyjny i/lub stóp zwrotu innych dostępnych dla firm możliwości inwestycyjnych.
W celu eliminacji wad metody NPV wynikających z nieprecyzyjności informacji autorka wprowadza jej wersje rozmyte poprzez rozmycie we wzorze (*) następujących parametrów:
a) n, w zawężeniu do realizacji liczb całkowitych bądź dowolnie,
b) CF_j (j = 0, ..., n), w zawężeniu do liczb nieujemnych lub niedodatnich bądź dowolnie,
c) r.

Badania operacyjne w analizie danych

Przykładem zastosowań metod rozmytych w zadaniach programowania liniowego jest praca Chanasa [4]. Jako współczynniki w funkcji celu autor przyjmuje liczby rozmyte. Wówczas wartość funkcji celu dla każdego rozwiązania dopuszczalnego, będąca oceną jego użyteczności, jest liczbą rozmytą. Formalnie zadanie to można sformułować w sposób następujący:

$$F(x) = \sum C_j x_j \rightarrow \max,$$

$$x \in X : \begin{cases} \sum a_{ij} x_j \le b_i & i = 1, \ldots, m, \\ x_j \ge 0 & j = 1, \ldots, n, \end{cases}$$

gdzie $C_j = (\underline{c}_j, \overline{c}_j, \alpha_j, \beta_j)$, $j = 1, \ldots, n$ są liczbami rozmytymi typu L-R.
Rozmyta postać funkcji celu jest konsekwencją faktu, że klasa liczb rozmytych typu L-R jest zamknięta ze względu na operację dodawania i mnożenia przez skalar.

Rozmyte modele taksonometryczne

Hierarchiczne skupianie zbiorów rozmytych

O dobrym dopasowaniu modeli rozmytych do procesów i zjawisk występujących w naukach pokrewnych ekonomii przekonuje praca Noworola [8]. Autor wykorzystuje hierarchiczną analizę skupień do konstrukcji testów psychometrycznych. Podobnie jak w przypadku klasycznej analizy hierarchicznej, podstawą skupiania obiektów bądź zmiennych jest tu macierz odległości, inaczej – macierz podobieństwa między nimi. Problem sprowadza się zatem do wyboru metryki właściwej danej sytuacji badawczej. W odniesieniu do konstruowania testów psychometrycznych autor zaleca stosowanie uogólnionej odległości euklidesowej oraz uogólnionej odległości Hamminga:

Uogólniona odległość Hamminga:

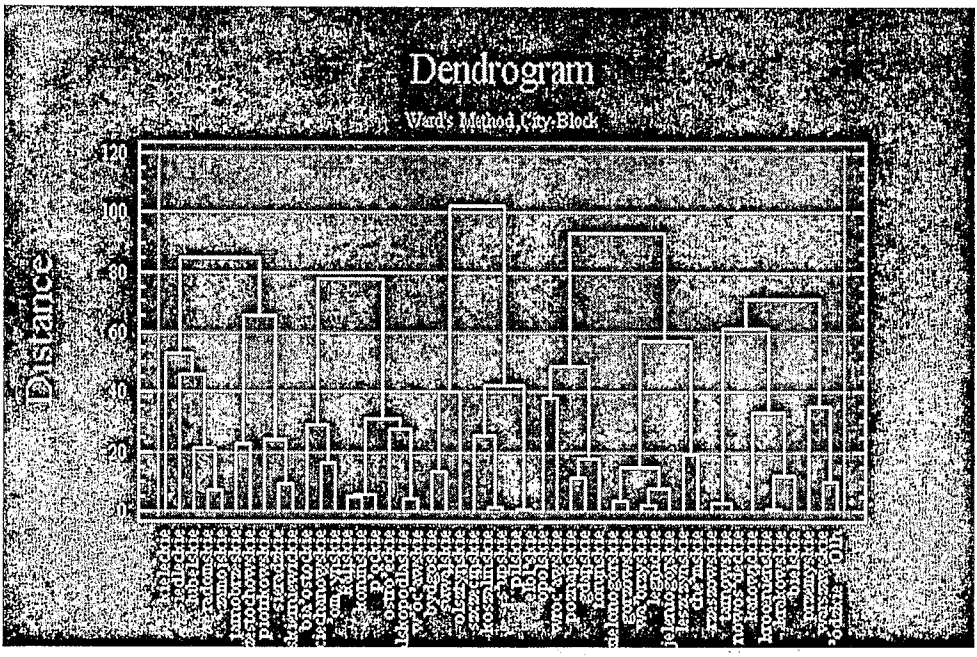
$$d_H(A,B) = \sum_{i=1}^{n} \left(MAX\left(\mu_A(x_i), \mu_B(x_i)\right) - MIN\left(\mu_A(x_i), \mu_B(x_i)\right)\right).$$

Uogólniona odległość euklidesowa:

$$d_H(A,B) = \left(\sum_{i=1}^{n} \left(MAX\left(\mu_A(x_i), \mu_B(x_i)\right) - MIN\left(\mu_A(x_i), \mu_B(x_i)\right)\right)^2 \right)^{\frac{1}{2}}.$$

Algorytmy klasyfikacji obiektów wynikające z analizy hierarchicznej można znaleźć w wersjach elektronicznych większości pakietów statystycznych. Użytkownik może wybrać spośród wielu dostępnych tam metod oraz odległości. Oprócz macierzy podobieństwa oraz specyfikacji skupień, można otrzymać dendrogram przedstawiający połączenia obiektów na różnych poziomach podobieństwa.

Na rysunku 2 przedstawiono przykładowy dendrogram otrzymany w wyniku zastosowania metody Warda z odległością miejską do klasyfikacji województw pod względem podobieństwa ich struktury agrarnej. Maksymalna liczba skupień – w analizowanym przypadku 5 – jest parametrem, który zadaje użytkownik.



Rysunek 1
Struktura powiązań województw na różnych poziomach podobieństwa ich struktury agrarnej

Relacja podobieństwa rozmytego

Nieco inną koncepcję rozmytych rozwiązań problemów taksonometrycznych prezentuje Ostasiewicz [9]. Tu pojęcie rozmytości zostało wprowadzone do formuły relacji podobieństwa, na bazie której dokonuje się klasyfikacji rodziny zbiorów. W odróżnieniu od klasycznej relacji podobieństwa, relacja podobieństwa rozmytego nie musi spełniać postulatów odległości. Wymaga się, by relacja ta spełniała warunek zwrotności oraz symetryczności, natomiast warunek przechodniości, właściwy każdej relacji równoważności, występuje w wersji zmodyfikowanej.

Formalny zapis funkcji definiującej relację podobieństwa rozmytego R oraz pierwszych dwóch warunków przedstawia się następująco:

$$\mu_R : \chi \times \chi \rightarrow [0, 1]; \text{ przy czym:}$$
$$\mu_R(a, a) = 1 \quad \text{oraz} \quad \mu_R(a, b) = \mu_R(b, a).$$

Złożenie R°R relacji rozmytej R określa się analogicznie do sposobu definiowania złożenia relacji klasycznej, przyjmując uogólnienia działań, oznaczone tutaj symbolami # oraz *:

$$\#: [0, 1]^2 \rightarrow [0, 1];$$
$$*: [0, 1]^2 \rightarrow [0, 1].$$

Zakłada się, że są to działania przemienne, łączne oraz niemalejące. Wymaga się ponadto, aby spełniały następujące warunki:

$$\min (a, b) \leq a \# b \leq \max (a, b);$$
$$0*0 = 0 ; 0*1 = 0 ; 1*1 = 1.$$

Relację $R°R = [r_{ij}]$ określa następujący wzór:

$$r_{ij} = \bigcup_{k=1}^{n} p_{ik} * q_{kj},$$

gdzie:

$$\bigcup_{k=1}^{n} a_i = a_1 \# a_2 \# ... \# a_n.$$

Wtedy warunek przechodniości relacji rozmytej R można zapisać następująco:

$$R \supseteq R°R.$$

W zależności od przyjętych definicji działań # oraz * otrzymuje się różne rodzaje przechodniości rozmytej.

Działania # i * , określone następująco: #: „+"; *: min, zastosowano do grupowania województw o podobnej strukturze wiekowej grubizny drzewostanów leśnych [2]. Mając macierz A:

$$A = \begin{bmatrix} a_{11}; & a_{12}; ... a_{1n} \\ a_{21}; a_{22}; ... a_{2n} \\ ..... \\ a_{m1}; a_{m2}; .. a_{mn} \end{bmatrix};$$

przedstawiającą udziały n zmiennych charakteryzujących m obiektów, relację podobieństwa rozmytego określa macierz $R = A \circ A^T$, gdzie T oznacza transpozycję macierzy. Zatem $r_{ij} = \sum_{t=1}^{n} \min(a_{it}, a_{tj})$, dla $i, j = 1, 2, m$.

Rozmyte metody wartościowania liniowego

Jedną z popularnych form przekazu informacji ekonomicznej są rankingi. Ranking, czyli uporządkowany pod względem jakości analizowanej cechy zbiór obiektów, służy z reguły do wyodrębnienia podzbioru obiektów najlepszych. Tu również można wprowadzić element rozmytości, definiując funkcję przynależności postaci (Cerioli, Zani):

$$f(o_i) = \begin{cases} 1, \text{ gdy } o_i \le c_1, \\ \dfrac{c_2 - o_i}{c_2 - c_1}, \text{ gdy } c_1 < o_i < c_2, \\ 0, \text{ gdy } o_i \ge c_2, \end{cases}$$

gdzie c_1 oznacza rangę progową, poniżej której przynależność do wyróżnionego zbioru jest gwarantowana, zaś c_2 – rangę progową, powyżej której przynależność ta jest niemożliwa. Formułę tę zastosowano w badaniach nad zróżnicowaniem skali inwestowania indywidualnych gospodarstw rolnych. Wyniki przedstawiono w formie tabelarycznej (tab. 1), zamieszczając zarówno wartości zmiennej syntetycznej, która stanowiła podstawę rankingu, jak i wartości funkcji przynależności.

Tak sformułowaną funkcję przynależności stosuje się również do ustalenia poziomu ryzyka ubóstwa. Adekwatność teorii zbiorów rozmytych w odniesieniu do pomiaru ubóstwa wynika z faktu, że przejście od stanu gorszej do lepszej sytuacji materialnej dokonuje się w sposób płynny. Dlatego też skonstruowano szereg charakterystyk rozmytych. Przykładowo, rozmyty indeks ubóstwa, FIP (ang. *Fuzzy Index of Poverty*):

$$FIP = \frac{1}{n} \sum_{i=1}^{n} \mu_P(i).$$

FIP interpretuje się jako procent populacji należącej w sensie rozmytym do podzbioru ubogich.

Tabela 1
Ranking województw ze względu na wielkość skali inwestowania indywidualnych gospodarstw rolnych

Województwo	Wartość zmiennej syntetycznej	Ranga	Wartość funkcji przynależności
kieleckie	8,873206	1	1
lubelskie	7,267246	2	1
zamojskie	6,241985	3	1
siedleckie	5,772944	4	0,956522
opolskie	5,618727	5	0,913043
kaliskie	5,174093	6	0,869565
tarnowskie	5,148303	7	0,826087
bydgoskie	5,126678	8	0,782609
tarnobrzeskie	5,084906	9	0,73913
rzeszowskie	5,003752	10	0,695652
poznańskie	4,942331	11	0,652174
radomskie	4,762446	12	0,608696
katowickie	4,724017	13	0,565217
białostockie	4,711953	14	0,521739
nowosądeckie	4,693512	15	0,478261
krakowskie	4,247282	16	0,434783
sieradzkie	4,018087	17	0,391304
konińskie	3,689062	18	0,347826
częstochowskie	3,609154	19	0,304348
piotrkowskie	3,58138	20	0,26087
ciechanowskie	3,579268	21	0,217391
toruńskie	3,465059	22	0,173913
łomżyńskie	3,462771	23	0,130435
bialskopodlaskie	3,358784	24	0,086957
płockie	3,356486	25	0,043478
gdańskie	3,158093	26	0
krośnieńskie	3,155732	27	0
olsztyńskie	3,134488	28	0
wrocławskie	3,10145	29	0
przemyskie	3,01644	30	0

cd. tab. 1

Województwo	Wartość zmiennej syntetycznej	Ranga	Wartość funkcji przynależności
bielskie	2,969265	31	0
leszczyńskie	2,796559	32	0
ostrołęckie	2,753406	33	0
skierniewickie	2,658571	34	0
pilskie	2,639397	35	0
włocławskie	2,549282	36	0
suwalskie	2,396454	37	0
chełmskie	2,200372	38	0
wałbrzyskie	1,857784	39	0
elbląskie	1,855929	40	0
zielonogórskie	1,766153	41	0
szczecińskie	1,703678	42	0
legnickie	1,519522	43	0
gorzowskie	1,359575	44	0
warszawskie	1,146369	45	0
koszalińskie	1,066384	46	0
słupskie	0,932102	47	0
jeleniogórskie	0,887614	48	0
łódzkie	0,387867	49	0

Literatura

1. BOGOCZ D., Regionalne zróżnicowanie struktury wiekowej grubizny krajowych drzewostanów leśnych, XXV Colloquium Biometryczne, AR w Lublinie, t. 27, 1997.
2. BOGOCZ D., *O rozmytych rozwiązaniach pewnych problemów taksonometrycznych*, Materiały z konferencji: Metody i zastosowania badań operacyjnych, Wrocław 1998.
3. BOGOCZ D., *Miejsce Polski w zbiorowości krajów Europy Zachodniej pod względem poziomu usług teleinformatycznych*, Zeszyty Naukowe AR w Krakowie, seria: Ekonomika, z. 27, 1999.
4. CHANAS S., *Zadanie programowania liniowego z rozmytymi współczynnikami w funkcji celu*, [w:] Metody i zastosowania badań operacyjnych, Praca zbiorowa pod red. T. Trzaskalika, Katowice 1998.

5. DRIANKOV D., HELLENDOORN H., REINFRANK M., *Wprowadzenie do sterowania rozmytego*, Wydawnictwa Naukowo-Techniczne, Warszawa 1996.
6. KUCHTA D., *Rozmyte metody wyboru i oceny inwestycji*, [w:] Metody i zastosowania badań operacyjnych, Praca zbiorowa pod red. T. Trzaskalika, Katowice 1998.
7. KUKUŁA K., *Ranking miejscowości turystycznych Ziemi Sądeckiej ze względu na stan bazy noclegowej*. Materiały z II Międzynarodowej Konferencji nt.: „Szanse i bariery rozwoju przedsiębiorczości w regionie Podkarpacia", 1998.
8. NOWOROL C., *Analiza skupień w badaniach empirycznych – Rozmyte modele hierarchiczne*, PWN, Warszawa 1989.
9. OSTASIEWICZ W., *Zastosowanie zbiorów rozmytych w ekonomii*, PWN, Warszawa 1986.

Some Applications of Fuzzy Set Theoryln Economic Research

Abstract

The paper presents some applications of fuzzy set theory based on examples from literature and own investigations.

The introduction contains some theoretical background with special attention given to different aspects of the definition and the characteristics of its properties.

In choosing empirical illustrations different fields of economic research are considered such as operation research, data analysis, taxonomy and linear valuation.

As the history of the theory starts with the theory of systems fuzzy controllers and their applications are shortly discussed.

Wacław Laskowski
Katedra Organizacji i Ekonomiki Konsumpcji SGGW

Ocena ilościowego spożycia żywności przez rodziny polskie z wykorzystaniem ich klasyfikacji przy użyciu sieci neuronowych

O ilości spożywanej żywności wiemy z danych GUS, tj. z danych bilansowych oraz z tzw. badań budżetów gospodarstw domowych, prowadzonych corocznie już od ponad 40. lat na losowanej corocznie próbie gospodarstw domowych. Konfrontacja znanych z tych źródeł przeciętnych wielkości spożycia z naukowymi zaleceniami żywnościowymi prowadzi do wniosku o dość zasadniczych między nimi rozbieżnościach, a to oznacza ryzyko chorób na tle wadliwego żywienia (Kowrygo 2000).

Dane z badań budżetów pozwalają dodatkowo stwierdzić, że istnieje duża indywidualna zmienność w spożyciu poszczególnych produktów, a nawet ich grup, przy tym skądinąd uzasadniona jest teza o istnieniu w odżywianiu się ludności względnie jednolitych stylów (wzorców) respektowanych konsekwentnie przez konkretne rodziny. Wzorce te warto by ocenić precyzyjnie ze zdrowotnego punktu widzenia, a w tym określić negatywne następstwa stosowanych racji pokarmowych. W tym celu przydatne byłoby wyłonienie takich grup gospodarstw domowych, które byłyby pod względem spożycia żywności względnie jednolite, a jednocześnie różniły się między sobą możliwie dużo, co upoważniałoby do uznania ich za reprezentujące odrębne style żywnościowe. Wydaje się, że sieci neuronowe mogą być bardzo pomocne w takiej klasyfikacji gospodarstw domowych.

Sieci neuronowe to w istocie urządzenia techniczne zbudowane na podstawie teorii formalnej o tej samej nazwie, której rozwój – zapoczątkowany oficjalnie w 1943 roku – był i jest nadal inspirowany biologiczną wiedzą człowieka o układach nerwowych [Hertz i inni 1995].

Sztuczne modele mózgu tworzone są ze znacznie mniejszej liczby neuronów jako podstawowych elementów składowych niż prawdziwy mózg. Sposób działania sieci neuronowych cechuje przede wszystkim to, że nie wymagają one algorytmu wykonania powierzanego im zadania. W to miejsce sieci wymagają wyuczenia (wytrenowania) przeprowadzanego na bazie przykładów. Proces uczenia sieci prowadzi do kształtowania wag połączeń pomiędzy poszcze-

gólnymi neuronami, które w ten sposób „przejmują" strukturę danych (sygnałów) i nabierają zdolności do generalizacji.

W koncepcjach neuronowych dostrzega się obiecujące możliwości zastosowań w wielu dziedzinach, a szczególnie w ekonomii, finansach, medycynie, geologii i fizyce, do realizacji różnych celów, a zwłaszcza tych, w których mamy do czynienia z klasyfikacją, przewidywaniem, optymalizacją, sterowaniem, rozpoznawaniem, filtrowaniem lub kojarzeniem danych (sygnałów) [Tadeusiewicz 1999].

Zasadniczym przełomem w rozwoju sieci neuronowych stała się możliwość ich symulowania przy użyciu współczesnych komputerów. Sprzyjająca jest duża ich szybkość i zasobność w pamięć operacyjną.

Jednocześnie coraz wyraźniej dokuczliwe stają się ograniczenia metod algorytmicznych i opartych na nich programach komputerowych. Dla fragmentarycznej tylko ilustracji uwarunkowań algorytmów można przywołać klasyczne procedury statystycznego opracowywania danych, które bardzo często wymagają spełniania wielu założeń, np. co do normalności rozkładu przypadków i liniowości związków pomiędzy analizowanymi cechami. Założenia te nie zawsze możliwe są do spełnienia, a przynajmniej wymagają sporej wiedzy już na wstępnym etapie analiz. Jest więc zapotrzebowanie na narzędzia, które byłyby bardziej praktyczne i uniwersalne.

Wiele wskazuje, że narzędziem takim mogą być sieci neuronowe. Ocenia się, że przy ich stosowaniu poziom wymaganej wiedzy teoretycznej do skutecznego zbudowania modelu aplikacyjnego jest znacznie mniejszy niż w przypadku tradycyjnych metod. Nie znaczy to oczywiście, że sieci są niezwykle prostym i zawsze skutecznym narzędziem, użytkownik musi bowiem dokonywać wyboru najlepszego typu neuronu, jak i określać architekturę (ich ułożenie) do zadania, jakie ma wykonać, wreszcie musi dysponować pewną wiedzą empiryczną pozwalającą na „wyuczenie sieci" [Tadeusiewicz 1999].

Sieć neuronowa typu Kohonena – wykorzystana przy prezentowanych analizach – jest siecią dwuwarstwową, w której podczas trenowania – w tym przypadku na całej zbiorowości – modelowane są wagi neuronów w taki sposób, by bardzo zbliżone do siebie przypadki reprezentował ten sam neuron, przy czym możliwe jest wprowadzenie zasady sąsiedztwa, pozwalającej na przejmowanie przez sąsiednie neurony przypadków podobnych. Zadaniem sieci Kohonena jest więc wykrycie „naturalnych" skupień przypadków ze względu na przyjęte cechy, co skłania do użycia określenia, że prowadzą one do podziału optymalnego.

Przedmiotem analizy są dane jednostkowe o spożyciu żywności z uwzględnieniem jej podziału na 13 produktów żywnościowych lub ich grup, spożyciu badanym przez GUS w 1998 roku w 31 756 gospodarstwach domowych traktowanych jako przypadki. Okresem uczestnictwa w badaniu jest jeden miesiąc, przy czym cała wylosowana próba składa się z podpróbek uczestniczących w badaniu w kolejnych miesiącach.

Analizuje się zatem charakterystyki poszczególnych gospodarstw domowych w postaci wielkości spożycia żywności przypadających (przeliczonych) na każdą osobę danego gospodarstwa. Są to wielkości w kg na miesiąc, z wyjątkiem jaj, które ewidencjonuje się w sztukach.

Metoda analizy wykorzystuje program Statistica z modułem Sieci Neuronowe formy StatSoft. Tworzono za pomocą tego modułu sieci neuronowe typu Kohonena o liczbie neuronów w warstwie wyjściowej od 40 do 150 przy zawsze 13 neuronach w warstwie wejściowej (tyle, ile zmiennych wejściowych). Trenowano z wykluczeniem sąsiedztwa, jak i z sąsiedztwem w promieniu 2 neuronów, na wszystkich danych; zwykle 30 epok (przebiegów) dawało najlepsze z możliwych ukształtowanie wag.

Wytrenowanymi sieciami klasyfikowano gospodarstwa domowe, czyli oznaczano je numerami neuronów zwyciężających w projekcji danego przypadku. Następnie oceniano, na ile powstałe skupienia gospodarstw przy tych samych neuronach wyjaśniają zmienność spożycia wyróżnionych produktów żywnościowych. Oceniano to przy użyciu stosunku korelacyjnego[1] ustalanego osobno dla każdego produktu, przy czym jako syntetyczne miary przyjęto stosunek średni, różnice korelacji poszczególnych produktów oraz liczbę uzyskanych przez daną sieć skupień (grup) i ich wielkości (tab. 1).

Sprawdzono też, w jakim stopniu wyjaśniana jest zmienność sumarycznego spożycia żywności. Ponieważ zmienna ta nie uczestniczy w modelu, to jej wyjaśnienie należy traktować jako efekt wtórny. Mimo to jej zmienność wyjaśniana jest też na zaskakująco wysokim poziomie (do 0,75). Wszystkie sieciowe grupowania okazały się dużo lepsze od klasycznego podziału na grupy społeczno-ekonomiczne, który daje wyjaśnienie zmienności spożycia przy wskaźniku średnim na poziomie 0,19 (tab. 1).

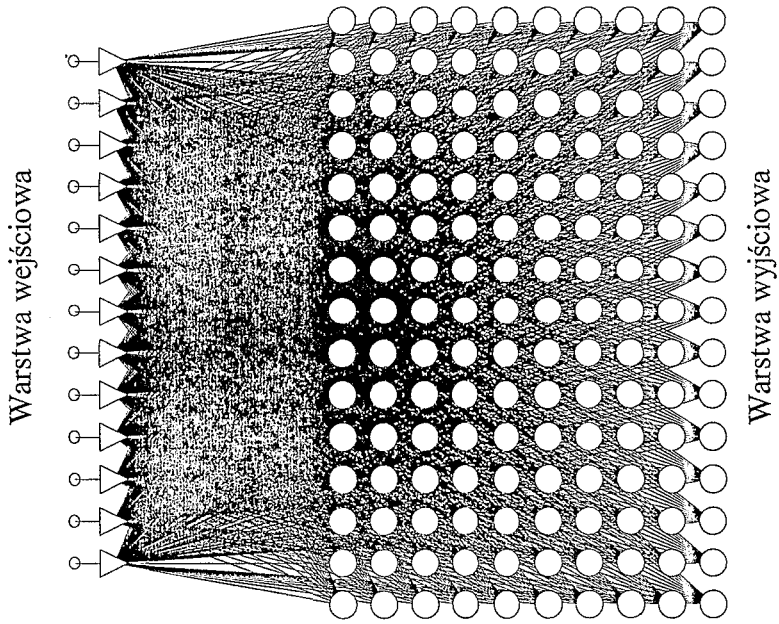
[1]Przyjmuje wielkości od 0 (brak korelacji) do 1 (związek funkcyjny).

Tabela 1
Korelacyjne charakterystyki klasyfikacji zaproponowanych przez różne sieci neuronowe i porównawczo klasyfikacji społeczno-ekonomicznej

Wyszczególnienie	Klasyfikacje gospodarstw domowych dokonywane przez sieci o liczbie neuronów:					Przynależność społeczno-ekonomiczna
	40	60	70	90 i sąsiedztwo	150	
Stosunki korelacji spożycia produktu w danych klasyfikacjach						
Zbożowe	0,50	0,51	0,52	0,60	0,54	0,30
Mięso i przetwory	0,53	0,65	0,65	0,77	0,63	0,17
Ryby i przetwory	0,34	0,67	0,65	0,76	0,65	0,13
Mleko	0,38	0,46	0,47	0,48	0,49	0,27
Jaja	0,42	0,59	0,61	0,77	0,64	0,20
Masło	0,17	0,25	0,41	0,35	0,45	0,16
Tłuszcze i oleje roślinne	0,78	0,85	0,79	0,89	0,85	0,21
Smalec, słonina	0,17	0,69	0,68	0,80	0,67	0,18
Owoce	0,66	0,66	0,64	0,79	0,65	0,22
Warzywa	0,43	0,68	0,49	0,78	0,69	0,17
Strączkowe	0,12	0,14	0,36	0,38	0,39	0,08
Ziemniaki	0,19	0,27	0,75	0,75	0,74	0,11
Cukier (też w wyrobach)	0,35	0,38	0,35	0,61	0,46	0,20
ŚREDNIO	0,39	0,52	0,57	0,67	0,60	0,19
Spożycie razem	0,58	0,64	0,73	0,79	0,75	0,29
Współczynnik zmienności korelacji (w %)	39	33	22	22	18	24
Liczba skupień obejmujących po co najmniej:						
10% zbiorowości	2	4	3	2	4	4
5% zbiorowości	4	5	5	5	4	7
1% zbiorowości	4	9	9	23	11	8
Łączna liczba grup	4	12	12	37	16	8

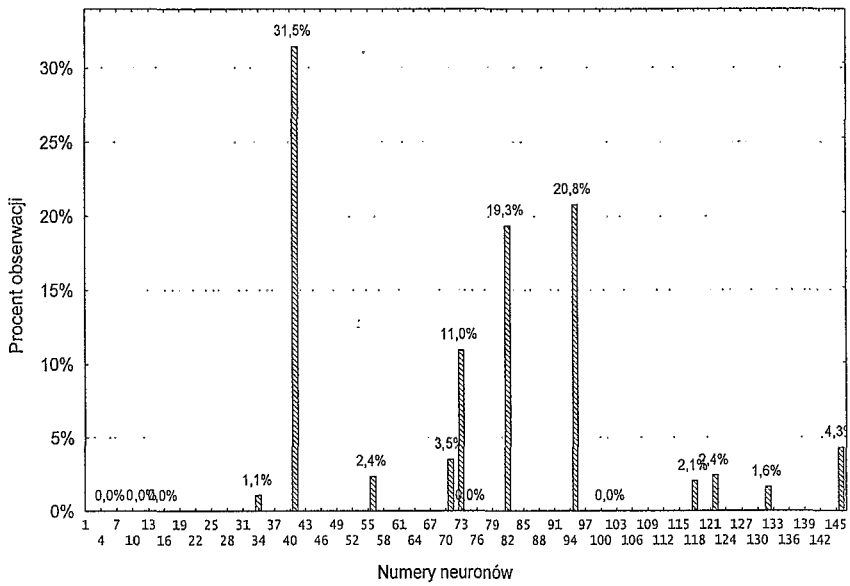
Źródło: Obliczenia własne na podstawie indywidualnych wyników badań budżetów gospodarstw domowych za 1998 rok oraz badań prowadzonych przez GUS.

Z punktu widzenia wygody dalszych analiz przydatniejsze okazało się korzystanie z klasyfikacji sieci Kohonena trenowanej bez sąsiedztwa. Sieci te dają mniej grup. Sieci o większej liczbie neuronów dawały rezultaty bardziej zachęcające. Do dalszych prezentacji wybrano zatem klasyfikację dokonaną przez sieć ze 150 neuronami, która dawała uśrednioną korelację na poziomie 0,60 przy najmniejszej skali zmienności tego wskaźnika w przekroju poszczególnych produktów, bo przy współczynniku tylko 18%, czyli przy stosunku korelacji od 0,39 w przypadku strączkowych do 0,85 w przypadku tłuszczów roślinnych. Korelacje te uznawane są jako wysokie. Za klasyfikacją tą przemawia też liczba uzyskanych grup i ich wielkości. Zachęcający jest podział z czterema względnie dużymi grupami obejmującymi po więcej niż 10% zbiorowości i siedmioma już wyraźnie mniejszymi, obejmującymi po od 1 do 5% zbiorowości (rys. 1 i 2).



Rysunek 1
Schemat sieci neuronowej Kohonena ze 150 neuronami

Od klasyfikacji wykonanej przez sieć ze 150 neuronami lepsza pod względem stopnia wyjaśniania ogólnej zmienności spożycia jest sieć z 90 neuronami trenowania z uwzględnianiem niewielkiego sąsiedztwa. Jednakże duża liczba zaproponowanych grup zasadniczo utrudnia i rozprasza ich głębszą analizę, należałoby je łączyć, co jednak prowadzi nieuchronnie do utraty siły wyjaśniania (tab. 1).



Rysunek 2
Histogram wielkości skupień utworzonych przez sieć Kohonena ze 150 neuronami

Podstawowe charakterystyki skupień gospodarstw domowych wyłonionych przez sieć Kohonena ze 150 neuronami zestawiono w tabeli 2. Poszczególne grupy rozmieszczono według malejącej ich wielkości, przy czym pominięto 5 bardzo małych grup obejmujących łącznie tylko 11 gospodarstw domowych. O tych gospodarstwach domowych powiedzieć można to, że odbiegają w sposób zasadniczy swymi charakterystykami od pozostałych. Wielkości ich dotyczące wzbudzają nawet podejrzenie o błąd albo są to bardzo szczególne przypadki, np. badane w okresie wesela, stąd o bardzo wysokich wielkościach spożycia. Z uwagi na to, że jest ich bardzo niewiele, pomija się je w dalszych prezentacjach. Niemniej jednak warto dostrzec zdolności sieci neuronowej do samodzielnego wyłonienia tego typu przypadków.

Pierwsza z prezentowanych grup jest najliczniejsza, obejmuje 31,5% zbiorowości, druga 20,8%, trzecia 19,3%, czwarta 11%, natomiast grupy od piątej do jedenastej obejmują już tylko po od niewiele ponad jeden do nieco ponad cztery procent, w sumie te ostatnie grupy obejmują 17,4% ogółu badanych. Można więc powiedzieć, że grupy od I do IV to grupy reprezentujące cztery względnie często spotykane style żywieniowe, pośród nich grupa I reprezentuje styl najczęstszy, dotyczący praktycznie co trzeciego gospodarstwa domowego. Pozostałe grupy reprezentują odmienne style, spotykane już rzadziej, a niektóre bardzo rzadko (tab. 2).

Tabela 2
Podstawowe charakterystyki grup gospodarstw domowych wg malejących ich wielkości i stosunku korelacji, będące rezultatem klasyfikacji wykonanej przez sieć Kohonena ze 150 neuronami

Wyszczególnienie		Razem	W tym skupienia (grupy gospodarstw domowych):										
			I	II	III	IV	V	VI	VII	VIII	IX	X	XI
Liczebność →		31 756	**9 992**	6 594	6 140	3 485	1 373	1 121	768	755	652	522	343
Liczebność w % →		100,0	**31,5**	20,8	19,3	11,0	4,3	3,5	2,4	2,4	2,1	1,6	1,1
Produkty spożywcze	Stosunek korelacji	Spożycie na 1 osobę, na miesiąc w kg (jaja w sztukach)											
Tłuszcze roślinne	0,85	1,3	**0,6**	1,5	1,0	2,5	1,1	1,9	1,7	1,4	1,7	5,0	1,6
Ziemniaki	0,74	9,2	**5,7**	6,7	8,9	9,2	7,4	11,6	18,8	11,8	8,2	14,5	132,0
Warzywa	0,69	6,5	**4,0**	4,3	7,2	7,4	8,8	9,1	31,2	7,7	6,9	11,7	12,2
Smalec, słonina	0,67	0,27	**0,15**	0,16	0,29	0,23	0,10	0,42	0,34	2,53	0,27	0,50	0,40
Ryby	0,65	0,44	**0,25**	0,36	0,40	0,51	0,59	0,62	0,43	0,54	3,11	0,96	0,41
Owoce	0,65	5,5	**3,8**	3,5	5,1	5,7	19,4	8,5	11,0	4,7	8,4	11,2	5,5
Jaja	0,64	17,4	**10,7**	12,9	22,8	20,7	16,6	47,6	23,0	21,4	21,6	36,1	18,1
Mięso i wyroby	0,63	6,0	**4,1**	4,8	7,5	7,1	5,6	12,2	7,7	10,3	8,5	10,9	7,0
Zbożowe	0,54	8,4	**6,3**	7,7	8,8	10,4	8,3	13,2	10,8	11,3	10,0	16,6	9,6
Mleko (też w wyrobach)	0,49	19,9	**15,1**	16,1	22,2	23,6	26,9	33,5	27,9	23,1	23,5	37,8	20,6
Cukier (też w wyrob.)	0,46	2,4	**1,6**	1,9	2,7	3,2	2,7	4,4	4,5	3,2	2,9	5,9	2,7
Masło	0,45	0,41	**0,36**	0,22	0,55	0,35	0,57	0,79	0,55	0,37	0,59	0,61	0,46
Strączkowe	0,39	0,11	**0,07**	0,08	0,13	0,13	0,13	0,16	0,27	0,20	0,18	0,20	0,15
Spożycie razem	0,75	61,4	**42,6**	48,0	66,1	71,4	82,5	99,1	116,4	78,5	75,5	117,9	193,5

Źródło: Jak w tabeli 1.

Zestawione w analizowanej tabeli przeciętne informacje o spożyciu poszczególnych produktów w poszczególnych grupach pozwalają owe style opisać i ocenić. Na samym początku zauważyć trzeba, że poszczególne grupy jaskrawo odróżnia sumaryczna wielkość spożywanej żywności na 1 osobę na miesiąc: od 42,6 kg w grupie I do 193,5 kg w grupie XI, przy czym ilość spożytej żywności systematycznie wzrasta w kolejnych grupach, z wyjątkiem VIII i IX. Prawidłowość dotycząca sumy spożytej żywności w dużym zakresie przenosi się na poszczególne produkty, zwłaszcza jeśli analizujemy grupy od I do IV, a więc te, które reprezentują najczęstsze style żywnościowe. W pozostałych grupach obserwujemy już częstsze zmiany proporcji w żywieniu lub pewne upodobania, np. do spożywania w większych ilościach masła zamiast tłuszczów roślinnych czy wyraźnie większe spożycie owoców i warzyw, ryb, smalcu albo ziemniaków.

Poszczególne grupy, które z uwagi na wysoki stopień wyjaśniania zmienności można uznać za reprezentujące występujące w rzeczywistości style żywnościowe, przedstawić można krótko w następujący sposób:

- **grupa I**, najliczniejsza, to osoby spożywające najmniej żywności; obserwujemy w tej grupie najniższe w przekroju wszystkich grup spożycie poszczególnych produktów, z wyjątkiem masła i owoców, których nieznacznie mniejsze spożycie jest tylko w grupie II,
- **grupa II**, tylko nieco wyższe niż w grupie I poziomy spożycia poszczególnych produktów, choć niektórych nieco wyraźniejsze (zbożowych, ryb, tłuszczów roślinnych), z wyjątkiem jednak jeszcze niższego niż w grupie I spożycia owoców i masła, przy czym to drugie jest kompensowane dosyć wysokim spożyciem tłuszczów roślinnych,
- **grupa III**, obserwujemy tutaj już wyraźnie wyższe poziomy spożycia wszystkich produktów, przy czym dotyczy to najbardziej spożycia jaj, smalcu, strączkowych, mięsa i warzyw; wielkości spożycia w tej grupie są najczęściej dosyć zbliżone do wielkości średnich w całej zbiorowości; warto ponadto zwrócić uwagę na umiarkowane spożycie tłuszczów roślinnych w tej grupie przy relatywnie dosyć wysokim poziomie spożycia masła, co jest odwrotnością sytuacji w II grupie,
- **grupa IV**, jest już stosunkowo mniej liczna, niemniej ponad 10% udziału świadczy o jeszcze znacznej popularności reprezentowanego przez nią stylu; wyróżnia go jeszcze wyższe niż we wszystkich poprzednich grupach spożycie żywności, na co składa się przede wszystkim wysokie spożycie tłuszczów roślinnych, a po nich wyrobów z dużą zawartością cukru, a także ryb, jaj i warzyw; ponadto, obserwujemy tutaj niskie spożycie masła przy bardzo wysokim poziomie spożycia tłuszczów roślinnych – można to zaliczyć do specyfiki tej grupy,

- **pozostałe grupy, czyli skupienia od V do XI,** są już wyraźnie mniejsze, niemniej jednak wyraźnie dystansują się od siebie i pozostałych, są grupami osób o wysokim odnotowanym poziomie spożywania żywności, przy czym w grupie V mamy osoby spożywające szczególnie dużo owoców, a także ryb; w grupie VI mamy osoby szczególnie obficie jedzące zbożowe, mięso, mleko, jaja, masło i cukier; w grupie VII mamy z kolei osoby spożywające najwięcej warzyw i strączkowych, a także dużo owoców, ziemniaków i cukru (są to więc osoby jedzące obficie, ale starające się uwzględniać zalecenia); w grupie VIII mamy osoby spożywające najwięcej smalcu oraz dużo strączkowych i mięsa; w grupie IX spotykamy osoby spożywające najwięcej ryb, przy jednocześnie sporym spożyciu owoców; w X grupie mamy osoby, które jedzą wyjątkowo obficie – podobnie jak w grupie VII – ale ze szczególną koncentracją na zbożowych, mleku, tłuszczach i cukrze; wreszcie XI grupę charakteryzują poziomy spożycia podobne do obserwowanych w grupie III, z wyjątkiem ziemniaków i warzyw, których wypada tu na 1 osobę bardzo dużo (można podejrzewać, że są to osoby badane w miesiącu gromadzenia zapasów, zwłaszcza ziemniaków).

Uzyskanie w miarę jednorodnych grup gospodarstw domowych pod względem spożycia żywności jest korzystne z punktu widzenia prowadzenia dalszej oceny, a przede wszystkim oceny zdrowotnej rejestrowanych u nich racji pokarmowych. Można bowiem stosunkowo precyzyjnie określać między innymi najbardziej niedoborowe w danej grupie składniki odżywcze i wskazywać negatywne tego następstwa. Wyniki takiej analizy mogą mieć duże znaczenie w edukacji żywieniowej i profilaktyce zdrowotnej. Ramy niniejszego artykułu nie pozwalają takiej oceny kontynuować.

Podsumowanie

Sieci neuronowe stanowią ciekawe i frapujące narzędzia badawcze. Mogą być bardzo przydatne w rozpoznawaniu danych, ich klasyfikacji, a także – jak wskazuje literatura – w wielu innych zastosowaniach. Nie wymagają rzeczywiście od badacza głębokiej wiedzy statystycznej czy ekonometrycznej, ale by rezultaty ich zastosowania były wiarygodne, trzeba mieć pewną wprawę w posługiwaniu się nimi. Teoria – jak na razie – nie dostarcza wystarczających reguł na potrzeby warsztatu badawczego.

Sieci neuronowe Kohonena otwierają drogę do klasyfikacji optymalnych, gdyż dają szansę na podziały przypadków przy bardzo wysokich współczynnikach wyjaśniania ogólnej zmienności uwzględnianych w modelu cech. W ten

sposób zwracają uwagę na ograniczoność poznawczą klasycznych podziałów. Analiza wyżej prezentowana wykazała, że możliwe jest uzyskanie tą drogą klasyfikacji gospodarstw domowych przy korelacji bardzo wysokiej i jednocześnie znacznie przewyższającej korelacje uzyskiwane przy typowych podziałach.

Warto ponadto zauważyć zdolności sieci neuronowych Kohonena do wyłaniania przypadków nietypowych, bardzo odbiegających. Może to mieć duże znaczenie w odnajdywaniu błędów lub eliminacji przypadków szczególnych, incydentalnych.

Przedstawiona analiza pozwoliła wyłonić grupy gospodarstw domowych, za którymi kryją się określone i zasadniczo odmienne style żywnościowe, które można poddać dalszej ocenie, a w tym ocenie zdrowotnej.

Literatura

1. HERTZ J., KROGH A., PALMER R., 1995: *Wstęp do teorii obliczeń neuronowych*, Wydawnictwa Naukowo-Techniczne, Warszawa.
2. KOWRYGO B.. 2000: *Studium wpływu gospodarki rynkowej na sferę żywności i żywienia w Polsce*, Wydawnictwo SGGW, Warszawa.
3. LASKOWSKI W., 2000: Sieci neuronowe w analizach statystycznych; *Wiadomości Statystyczne* 7, GUS, Warszawa.
4. OSOWSKI S., 1996: *Sieci neuronowe*, Wydawnictwa Naukowo-Techniczne, Warszawa.
5. TADEUSIEWICZ R., 1993: *Sieci neuronowe*, Akademicka Oficyna Wydawnicza RM, Warszawa.

Quantitative analysis of food consumption by Polish families with the utilization of their classification with the use of neural networks

Abstract

Data from surveys carried out by GUS permit to state that there is large individual variability in consumption of food by the Polish population. However, traditional partitioning of households, for example, according to income or dwelling place does not fully explain this variability. It was assumed that neural networks as a new research tools can shade more light is respect to consumption of food in concentrated households.

The use of neural networks of the Kohonen type enabled the separation of 4 basic concentrations of households differing in the level of food consumption as well as in the contribution of various products. This partitioning explained at a high level the variability in food consumption (the average correlations coefficient of groups of products examined reached the level of 0.6).

Moreover this partitioning gave the opportunity to further analyze relatively homogenous households that clarity differ with each other.

Ludosław Drelichowski
Katedra Ekonomiki Rolnictwa i Informatyki
Akademii Techniczno-Rolniczej w Bydgoszczy

Zastosowanie metod optymalizacyjnych w systemach logistyki agrobiznesu

Wprowadzenie

Procesy występujące w zaawansowanej gospodarce rynkowej prowadzą do standaryzacji nowoczesnych technologii wytwarzania i organizacji produkcji dostępnych na rynku dóbr inwestycyjnych. Oznacza to, iż poszukiwanie źródeł przewagi konkurencyjnej poprzez zastosowanie nowoczesnych technik i technologii wytwarzania nie stanowi gwarancji sukcesu. Plastycznie prezentowany w „łańcuchu wartości Portera"[1] czynnik przewagi konkurencyjnej jako pochodna rozwiązań głównych i pomocniczych funkcji zarządzania jest coraz trudniejszy do uzyskania we współczesnych uwarunkowaniach.

Interesującym źródłem rezerw w tym zakresie stały się w ostatnich 20 latach rewolucyjne zmiany procesów logistycznych. Przegląd podstawowych problemów logistyki w ujęciu rozwiązań merytorycznych i systemów informacyjnych zawiera praca Abta[2]. Objęły one przede wszystkim rozwiązania metod zaopatrzenia w surowce i materiały, dystrybucję wyrobów, a często również różne fazy procesów wytwarzania. Istota tych rozwiązań polega na outsourcingu tych funkcji w przedsiębiorstwie, co oznacza ich przekazanie wyspecjalizowanym przedsiębiorstwom. Taka procedura powoduje standaryzację procesów zaopatrzenia i dystrybucji, co sprzyja zastosowaniu metod optymalizacyjnych.

Takie przesłanki powodowały, że standaryzacja procedur dystrybucyjnych, powszechne stosowanie kodów kreskowych lub identyfikatorów produktu na kartach magnetycznych, pozwalała precyzyjnie identyfikować obrót towarowy i osiągać korzyści pochodzące ze wzrostu sprawności procesów zaopatrzenia i dystrybucji. Wśród nośników postępu czołowe miejsce w oprogramowaniu specjalistycznym tego typu systemów zajmują modele optymalizacyjne.

---

[1]Porter M.E.: *Strategie konkurencji.* PWE, Warszawa 1994.

[2]Abt S.: *Zarządzanie logistyczne w przedsiębiorstwie.* PWE, Warszawa 1998.

W prezentowanej pracy przedstawiono dwa z szerzej stosowanych systemów specjalistycznych z tego zakresu, tj. amerykański MANUGISTICS oraz francuski GOLD.

System GOLD został zaprezentowany w szerszym zakresie oferowanych w nim funkcji użytkowych i struktury systemu, podczas gdy system MANUGISTICS został zastosowany w przedsiębiorstwie STAR-FOODS będącym dużym producentem chipsów oraz wyrobów cukierniczych o długim okresie trwałości. Ważnym i weryfikowalnym w praktyce elementem w zastosowaniu tego typu systemów jest realność uzyskania zmniejszenia kosztów logistyki w granicach 15–25%, osiąganego u wielu użytkowników. Źródłem tych sukcesów jest właśnie zastosowanie szerokiej palety modeli optymalizacyjnych, które zastały w tych pakietach zastosowane do tworzenia codziennych wariantów marszrut dystrybucyjnych w tego typu przedsiębiorstwach.

Determinanty zastosowań modeli optymalizacyjnych w systemach logistycznych

Omawiana we wprowadzeniu standaryzacja rozwiązań organizacyjnych, przedsiębiorstw logistycznych oraz procedur postępowania niezależnych od specyfiki branż, których obsługę firma logistyczna realizuje, znakomicie upraszcza zastosowanie modeli optymalizacyjnych do wspomagania procesów decyzyjnych. Otwarcie na rynek usług logistycznych adresowanych do możliwie szerokiej grupy odbiorców powoduje, iż projektowanie systemów informacyjnych dla tego typu firm było przystosowane do rozwiązywania uniwersalnych i specjalistycznych funkcji logistycznych w ramach opracowanych algorytmów. W rachubę wchodzi tutaj rozwiązywanie problemów zaopatrzenia materiałowego i dystrybucji wyrobów poprzez tworzenie łańcuchów logistycznych obsługujących kontrahentów w ramach funkcji zleconych bądź też korzystanie z sieci logistyczno-dystrybucyjnej firmy specjalistycznej. Specyficzny przykład tego typu funkcji usługowych stanowić mogą sieci hipermarketów[3], które aktywnie oddziałują na rozwiązania organizacyjne dostawców, podejmujących współpracę z hipermarketami.

Do zakresu podejmowanych w outsourcingu usług logistycznych włączana jest coraz częściej obsługa zaopatrzenia produkcji w niezbędne materiały, często z realizacją obsługi w systemie JIT (just in time), wymagająca zastosowania

[3]Drelichowski L.: *Elementy teorii i praktyki zarządzania z technikami informacyjnymi w przedsiębiorstwie*. ATR, Bydgoszcz 2000.

interakcyjnych systemów informatycznych oraz odpowiednio zlokalizowanego zaplecza magazynowego. Obsługa tak złożonych procesów koordynacji dostaw, połączona z intensywnością realizacji transakcji magazynowych i fakturowania, stwarza możliwość uzyskiwania istotnych oszczędności, wynikających ze scalenia rozwiązań logistycznych kilkudziesięciu zwykle czy kilkuset firm. Tak duże strumienie informacyjno-zasileniowe zachęcają również do wprowadzania rozwiązań optymalizacyjnych, czemu sprzyja wysoki poziom standaryzacji procedur informacyjno-decyzyjnych.

Powyższe przyczyny powodowały, iż twórcy oprogramowania standardowego dla przedsiębiorstw logistycznych uwzględniali możliwość zastosowania modeli optymalizacyjnych. Istotnym czynnikiem, decydującym o skali zastosowań oraz efektywności tych rozwiązań, była decyzja o wykorzystaniu tych modeli do wspomagania decyzji operacyjnych. Ściślej, było to rozwiązanie z zakresu automatyzacji procesów decyzyjnych połączone z wystawianiem dokumentów źródłowych, które stanowią podstawę do wystawiania dokumentów rozliczeniowych (dokumenty magazynowe, karty drogowe, harmonogramy dostaw). Często mogą to być decyzje przesądzające o technicznym rozmieszczeniu materiałów, decydujące o czasie dostępu do półki z danym asortymentem w magazynie wysokiego składowania.

W świetle powyższych rozważań można stwierdzić, że specyficzne uwarunkowania działalności przedsiębiorstw logistycznych oraz skala realizowanych transakcji w powiązaniu ze standaryzacją procedur informacyjno-decyzyjnych sprzyjają rozwojowi zastosowań modeli optymalizacyjnych w stopniu przekraczającym jakiekolwiek doświadczenia uzyskiwane w innego typu przedsiębiorstwach. Taki stan sprzyja osiąganiu wysokich efektów działalności przedsiębiorstw logistycznych oraz uzyskiwaniu przewagi konkurencyjnej u przedsiębiorstw zlecających działalność logistyczną wyspecjalizowanym firmom.

Czynniki umożliwiające standaryzację procedur warunkujących łatwość zastosowania modeli optymalizacyjnych do podejmowania decyzji operacyjnych

Omówione w poprzednich rozdziałach uwarunkowania efektywnego działania przedsiębiorstw logistycznych wskazują obszary zastosowań modeli optymalizacyjnych, których wykorzystanie poprzedzało doskonalenie rozwiązań organizacji zarządzania i systemów informatycznych. Obszary procesów informacyjno-decyzyjnych wskazane do priorytetowego wspomagania zasto-

sowaniem modeli optymalizacyjnych dotyczą usprawnień systemu obsługi klienta lub obniżenia kosztów dokonywania zakupów i dystrybucji wyrobów dla grup kontrahentów. Szerszą prezentację teoretycznych aspektów podejmowania decyzji logistycznych w przedsiębiorstwie prezentuje Kisperska-Moroń[4], wskazując przesłanki uzasadniające występowanie możliwości zastosowań modeli optymalizacyjnych.

Nowe podejście do rozwiązań problemów logistyki przedsiębiorstw kształtowane było przez dwie specyficzne tendencje rozwojowe, związane z jednej strony ze zjawiskami koncentracji i globalizacji przedsiębiorstw, z drugiej zaś z działalnością wyspecjalizowanych firm logistycznych. W obydwu rozpatrywanych przypadkach kierunki działań i logika usprawniania procesów logistycznych prowadziły do tworzenia zbliżonych algorytmów oraz procedur postępowania. Skala działalności przedstawicieli obydwu grup przedsiębiorstw była realizowana lub zmierzała do realizacji na rynku globalnym, co powodowało unifikację potrzeb obydwu ważnych grup użytkowników tworzącego się nowoczesnego oprogramowania dla firm logistycznych. Można stwierdzić, że kierunki unowocześnienia logistyki przedsiębiorstw inspirowane były komplementarnymi potrzebami umacniających się na rynku firm globalnych oraz rozwijających się w szybkim tempie firm logistycznych. Oznaczało to możliwość poszerzenia potencjalnej listy beneficjentów postępu dokonującego się w logistyce, nie tylko do grona wymienionych wyżej dwu grup firm, ale również tej grupy przedsiębiorstw, które zdecydują się na realizację obsługi procesów logistycznych w formie outsourcingu.

Czynnikiem decydującym o wielostronnych sukcesach rozwoju systemów logistycznych stało się radykalnie nowe spojrzenie na role tych procesów w systemie organizacji zarządzania przedsiębiorstwem oraz postrzeganie tych procesów jako potencjalnych źródeł uzyskiwania przewagi konkurencyjnej. Dokonywany postęp w nowym postrzeganiu funkcji logistycznych w przedsiębiorstwach, standaryzacja i unifikacja procedur w nich stosowanych, oderwanie od tak często wyodrębnianego postrzegania branżowego, stanowiły doskonały punkt wyjścia do opracowania systemów informacyjnych nowej generacji z tego zakresu.

Kompleksowa, tworzona w czasie rzeczywistym baza danych systemu zapewniała możliwość udostępnienia bądź tworzenia parametrów niezbędnych w procesach modelowania. Stanowiło to ostatni istotny element umożliwiający opracowanie modeli optymalizacyjnych tworzonych w celu automatyzacji pro-

---

[4]Kisperska-Moroń D.: *Podstawy podejmowania decyzji logistycznych w przedsiębiorstwie*. AE, Katowice 1995.

cedur decyzyjnych, służących do podejmowania decyzji operacyjnych. Na podstawie tych decyzji dokonywane jest tworzenie dokumentów transakcyjnych, opracowywanych w postaci niezbędnej do kompleksowej obsługi procesów logistycznych.

Przedstawiona w tym rozdziale analiza przesłanek decydujących o realności szans w zakresie zastosowań modeli optymalizacyjnych w procesach logistycznych w stopniu przekraczającym dotychczasowe doświadczenia w zarządzaniu przedsiębiorstwami wskazała, że punktem wyjścia do sukcesu było radykalnie nowe podejście do organizacji procesów magazynowych i zaopatrzeniowo-dystrybucyjnych w przedsiębiorstwach.

Obszary zastosowań modeli optymalizacyjnych w systemach logistycznych

Zakres zastosowań modeli optymalizacyjnych w systemach logistycznych sprecyzowano na podstawie standardów wypracowanych w ramach systemu GOLD, a których funkcje użytkowe zostaną krótko scharakteryzowane w tym rozdziale.

Pierwsze dwa typy modeli optymalizacyjnych dotyczą problemów optymalnego rozmieszczenia zapasów na podstawie zawartej w bazie danych statystyki obrotów magazynowych i lokalizacji odbiorców poprzez:

- optymalizację składowania towarów wg adresów magazynowych,
- optymalizację struktury sieci dystrybucyjnej.

Kolejny model optymalizacyjny może być wykorzystywany w magazynach płaskich bądź magazynach wysokiego składowania, uzależniając miejsce składowania od intensywności obrotów danym asortymentem materiału. Celem jest minimalizowanie czasu trwania obsługi technicznej transakcji, który uzależniony jest od sposobu rozmieszczenia materiałów w zależności od intensywności jego uczestniczenia w obrotach przez optymalizację tras wózków widłowych w magazynach.

Następny problem dotyczy klasycznego modelu optymalizacji tras transportu samochodowego, sporządzanego okresowo na podstawie statystyki obrotów realizowanych w ramach modułu optymalizacji tras transportu samochodowego.

Następny model optymalizacyjny dotyczy ewidencji i optymalizacji kosztów wg zdefiniowanych miejsc ich powstawania w strukturze logistycznej. Jego efektywne działanie możliwe jest po zrealizowaniu modułu ewidencyjnego,

który zapewnia odpowiednią bazę parametrów do modułu optymalizacyjnego.

Kolejnym modelem zawierającym procedury ekonometryczne i heurystyczne jest prognozowanie sprzedaży, które w przypadku systemu MANUGISTICS doprowadzone jest do poziomu opracowywania operacyjnych planów produkcji przedsiębiorstwa.

Inny nieco charakter ma oprogramowanie dotyczące tworzenia wielokryterialnych analiz statystycznych sprzedaży, zakupu i stanów asortymentu, które rzutuje na jakość materiałów analitycznych wspomagających podejmowanie decyzji w przedsiębiorstwie. Standaryzacja serwisu informacyjnego z tego zakresu stanowi ważny element doskonalenia analizy stanu i doskonalenia podejmowania decyzji logistycznych w przedsiębiorstwach.

Kolejne procedury zawierające modele optymalizacyjne dotyczą optymalizacji zakupów (uzupełnień zapasów), które często korzystają z metod ABC oraz XYZ, szerzej prezentowanych w pracy Skowronka i Sariusza-Wolskiego[5].

Najbardziej skomplikowane zadania optymalizacyjne realizują modele wielokryterialnej optymalizacji terminów, kosztów dostaw i ustalania marszrut transportowych, wykorzystywane do operacyjnego rozdziału zadań dystrybucyjnych uwzględniających priorytety dla terminów, kosztów i sekwencji realizacyjnych (marszrut).

Powyższe grupy modeli stanowią najczęściej wykorzystywany w pakietach programowych systemów logistycznych standard systemów wspomagania decyzji. Z niewielkim ryzykiem popełnienia błędu można stwierdzić, że to właśnie modele optymalizacyjne zawarte w konstrukcji Pakietów Programowych Logistyki decydują o powszechnie weryfikowalnych efektach ekonomicznych wynikających z ich zastosowania.

Analiza funkcji użytkowych pakietu standardowego systemu logistycznego GOLD (Global Optimised Logistics and Distribution)

Funkcje użytkowe pakietu logistycznego GOLD realizują kompleksową obsługę procesów transakcyjnych, zapewniając integrację procesów dotyczących obsługi klienta oraz ewidencji obrotu magazynowego. Procesy transakcyjne dostarczają zasileń on-line bazy danych, pozwalających zlokalizować w czasie i przestrzeni przepływy zasileniowe i odpowiadające im strumienie

[5]Skowronek M., Sariusz-Wolski Z.: *Logistyka w przedsiębiorstwie*. PWE, Warszawa 1995.

informacyjne. Precyzyjna identyfikacja operacji przemieszczania materiałów z identyfikacją podmiotów tego obrotu zapewnia uzyskiwanie precyzyjnych parametrów niezbędnych do automatycznego generowania parametrów do modeli optymalizacyjnych wykorzystywanych do optymalizacji decyzji operacyjnych.

Funkcje zarządzania zasilane przez system dotyczą:
- obsługi transakcji magazynowych,
- polityki cenowej,
- zamówień klientów (obsługi indywidualnej klientów i promocji),
- fakturowania oraz kontroli realizacji płatności,
- bonifikat, upustów oraz ustalania harmonogramów płatności,
- obsługi inwentaryzacji;
- identyfikacji asortymentu i parametrów jego jakości.

Główne funkcje zintegrowanego zaopatrzenia polegają na automatyzacji procedur:
- zarządzania danymi głównymi,
- uzupełniania zapasów,
- polityki cenowej,
- zarządzania magazynem,
- zarządzania sklepami,
- emisji statystyki obrotów i wydruku raportów,
- kontroli faktur zakupu;
- obsługi promocji i akcji specjalnych,
- kompleksowego zarządzania rabatami zwrotnymi (premie na koniec roku),
- określania polityki handlowej,
- przeprowadzania operacji specjalnych (promocje itp.),
- kontroli zarządzania,
- komunikacji z partnerami/dostawcami.

Omawiane wyżej funkcje realizowane są niezależnie od omówionych poprzednio modułów zawierających modele optymalizacyjne. Zakres informacji dotyczących struktury systemów logistycznych zawartych w tej pracy z konieczności ogranicza się tylko do zasygnalizowania najważniejszych elementów, stanowiących moduły wykonawcze systemu. Rozwinięcie tych zagadnień radykalnie przekraczałoby dopuszczalną objętość niniejszego opracowania, zainteresowanych czytelników odsyłam do cytowanych pozycji literatury.

Ocena doświadczeń zastosowania pakietu logistycznego MANUGISTICS w STAR-FOODS SA w Tomaszowie Mazowieckim

System MANUGISTICS należy do jednego z najszerzej stosowanych w świecie rozwiązań oprogramowania dla obsługi procesów logistycznych. Złożoność procesów logistycznych występujących w przedsiębiorstwie STAR-FOODS SA z Tomaszowa Mazowieckiego i skala działalności uzasadniała zastosowanie tego kosztownego systemu do doskonalenia procesów logistycznych. Przedsiębiorstwo to należy do największych w Polsce producentów chipsów, popcornu i ciast długiego przechowywania (rogale z czekoladą), prowadzi też hurtownie importowanych artykułów spożywczych (np. orzeszki ziemne, bakalie itp.).

Skalę działalności przedsiębiorstwa przybliżyć może występowanie 16. hurtowni regionalnych zwanych magazynami-depo z siecią bezpośredniej sprzedaży oraz strumieni faktur liczących 3 do 5 tysięcy dokumentów dziennie. System analizy dystrybucji towarów i wyrobów dostarcza niezbędnych informacji do tworzenia automatycznych wariantów planu produkcji, tworzonych krocząco w cyklu tygodniowym i miesięcznym.

Uzyskiwane z systemu MANUGISTICS propozycje szczegółowych dyspozycji transportowych dla określonych wyrobów i planów sprzedaży zapewniają zmniejszenie kosztów realizacji dostaw w granicach od 10 do 15 procent. Zaletą omawianego systemu jest jego elastyczność w stosunku do wymagań systemu zasilającego danymi transakcyjnymi jego bazę. Wprawdzie osiągany zakres zaawansowanych funkcji optymalizacyjnych zależy od możliwości identyfikacyjnych bazy kodowej systemu ewidencyjnego, jednak konstrukcja systemu MANUGISTICS zapewnia występowanie dużej skali adaptacji determinowanej strukturą bazy kodowej.

Omawiany tutaj przykład zastosowań systemu logistycznego w globalnym przedsiębiorstwie agrobiznesu (znaczna sprzedaż eksportowa) wskazuje na wysoką użyteczność tego typu rozwiązań w innych dużych przedsiębiorstwach, co prawdopodobnie będzie determinowało przewagę konkurencyjną tych organizacji. Rozwój tego rodzaju zastosowań w polskich przedsiębiorstwach jest w dużym stopniu ograniczany wysokimi kosztami opłat licencyjnych (0,5–1,5 mln USD), stosunkowo małą skalą działalności polskich firm logistycznych oraz nie dość nowoczesnymi rozwiązaniami organizacji logistyki w dużych przedsiębiorstwach. Uzyskanie bardziej precyzyjnej oceny przyczyn istniejącego stanu wymagałoby przeprowadzenia dodatkowych badań.

Wnioski

1. Wysokie koszty opracowania nowych rozwiązań technologii wytwarzania i organizacji produkcji wymagają ich standaryzacji w skali globalnej. Oznacza to, iż punkt ciężkości w dążeniu do osiągnięcia przewagi konkurencyjnej w większym stopniu uzależniony jest od zastosowanych rozwiązań logistycznych.

2. Jednym z upowszechniających się rozwiązań logistyki przedsiębiorstw jest stosowanie koncepcji outsourcingowych.

3. Rezultatem rozwiązań sprecyzowanych we wniosku 2 jest dynamiczny rozwój dużych, działających globalnie firm logistycznych.

4. Tworzenie specjalistycznych firm logistycznych oraz doskonalenie rozwiązań logistycznych przedsiębiorstw globalnych doprowadziło do standaryzacji procedur oraz procesów w nich stosowanych.

5. W rezultacie procesów sformułowanych we wnioskach 3 i 4 stworzono specjalistyczne pakiety programów logistycznych, w których możliwe było zastosowanie szerokiej gamy modeli optymalizacyjnych, wykorzystywanych do automatyzacji decyzji operacyjnych.

6. Rozwój zastosowań tego typu pakietów specjalistycznego oprogramowania w dużych przedsiębiorstwach może stać się jednym ze źródeł obniżki kosztów.

Literatura

1. ABT S.: *Zarządzanie logistyczne w przedsiębiorstwie*. PWE, Warszawa 1998.
2. DRELICHOWSKI L.: *Elementy teorii i praktyki zarządzania z technikami informacyjnymi w przedsiębiorstwie*. ATR, Bydgoszcz 2000.
3. KISPERSKA-MOROŃ D.: *Podstawy podejmowania decyzji logistycznych w przedsiębiorstwie*. AE, Katowice 1995.
4. PORTER M.E.: *Strategie konkurencji*. PWE, Warszawa 1994.
5. SKOWRONEK M., SARIUSZ-WOLSKI Z.: *Logistyka w przedsiębiorstwie*. PWE, Warszawa 1995.

Application of optimisation methods in logistic systems of agribusiness

Abstract

In the paper were presented circumstances decided about a dynamic development of new methods of logistics management in companies. Introduction of new logistic solutions was realised simultaneously with updating new specific standard software supporting dissemination of modern organisational concepts.

In the work were described also optimisation models used in packages as procedures of making of operational decisions, what is unique standard applied in management information systems.

Sylwester Smolik
Katedra Zastosowań Matematyki SGGW
Piotr Łukasiewicz
Katedra Ekonometrii i Informatyki SGGW

Opis zjawisk okresowych z monotonicznie zmieniającą się amplitudą

Wprowadzenie

W badaniach ekonomicznych, technicznych i przyrodniczych bardzo często uzyskujemy wartości badanej cechy czy zjawiska w różnych momentach czasu. Tego rodzaju dane, uporządkowane w ciągi (t, y_t), dla $t = 1, 2, ..., n$, nazywamy szeregiem czasowym. Takie zbiory danych prezentujemy za pomocą tabeli lub wykresu, jako punkty płaszczyzny połączone odcinkami linii prostej. Już z wizualnej obserwacji wspomnianego wykresu często wnioskujemy, że badane zjawisko podlega typowym prawidłowościom, których wykrycie i opis ilościowy są celem analizy szeregów czasowych. Patrząc na rysunek 1, obrazujący skup mleka w Polsce w latach 1970–1979, spostrzegamy, że prezentowane zjawisko podlega wyraźnej tendencji wzrostowej (trend) oraz że w kolejnych latach mają miejsce podobne kształtem wahania okresowe, ale o rosnącej z latami amplitudzie. Są one obciążone pewnym błędem pomiaru. W teorii ekonomicznych szeregów czasowych nazywa się je wahaniami przypadkowymi. Powszechnie przyjęta metoda analizy szeregów czasowych polega na rozłożeniu ich wartości na trzy składniki – trend, składnik okresowy (w szczególności sezonowy) i składnik losowy. Postępowanie to tłumaczy się tym, że wymienione wyżej składniki szeregu czasowego są wynikiem działania trzech wyłączających się kompleksów przyczyn kształtujących ich nasilenie. Zgodnie z tym będziemy analizowali takie zjawiska, które można rozłożyć na wymienione wyżej trzy składniki:

$$y_t = f(t) + z(t) + \varepsilon_t \qquad (t = 1, 2, ..., n). \tag{1}$$

Wyodrębnienie w badanym zjawisku tendencji rozwojowej przez eliminację jego wahań okresowych i przypadkowych nazywamy wyrównywaniem lub wygładzaniem szeregu czasowego. Zwykle wyróżnia się dwie metody wygładzania: mechaniczną i analityczną. W mechanicznej metodzie używa się średnich ruchomych. Liczba składników średniej ruchomej powinna być odpowiednio dobrana. W analitycznej metodzie wygładzania tendencję rozwojową badanego zjawiska opisujemy funkcją gładką, a jej parametry estymujemy metodą najmniejszych kwadratów. Naszym zdaniem, parametry funkcji trendu przy obecności wahań okresowych wygodnie jest estymować metodą średnich, analogicznie jak to czyni się w metodzie mechanicznego wygładzania. W ten sposób nie tylko zachowujemy taką samą metodę przy wyznaczaniu tendencji rozwojowej, ale tylko ta metoda zapewnia sprowadzenie oscylacji przekształconego szeregu czasowego wokół zera. Pozwala to na zmniejszenie o jeden liczby parametrów koniecznych do opisu okresowości zjawiska. Dokładniejsze objaśnienia przytoczymy w czwartej części niniejszego artykułu.

Estymacja parametrów modelu wahań okresowych

Zakładamy, że dany jest ciąg punktów empirycznych (t, y_t) dla $t = 1, 2, ..., n$, obrazujących przebieg badanego zjawiska. Ocenę tendencji rozwojowej $f(t)$ można wyznaczyć różnymi metodami. Wygodnie jest estymować parametry funkcji $f(t)$ wspomnianą wcześniej metodą średnich, wówczas, zgodnie z jej definicją, obliczamy szukane wartości ocen parametrów regresji z warunku zerowania się sumy odchyleń. Pozostały składnik danego szeregu czasowego $z_t = y_t - \hat{f}(t) - \varepsilon_t$ opiszemy harmoniką postaci:

$$z_t = (A + B \cdot t) \cdot \sin(\omega\, t + \theta) - \varepsilon_t, \text{ gdzie } \omega = 2\pi/T \quad (t = 1, 2, ..., n), \qquad (2)$$

gdzie: ω – częstość (prędkość kątowa), T – okres badanego zjawiska, θ – jego faza początkowa, A i B to poszukiwane stałe, z których buduje się amplitudę harmoniki. Może być ona stała lub zmieniać się monotonicznie. Parametry modelu (2) szacujemy metodą najmniejszych kwadratów. W tym celu poszukujemy minimum następującej funkcji:

$$F(A,B,\theta,\omega) = \sum_{t=1}^{n} \varepsilon_t^2 = \sum_{t=1}^{n} \left[z_t - (A + B \cdot t)\sin(\omega t + \theta) \right]^2 = \text{minimum}. \qquad (3)$$

Obliczamy pochodne cząstkowe funkcji (3):

$$\begin{aligned}
\frac{\partial S}{\partial A} &= -2\cdot\sum_{t=1}^{n}\left[z_t-(A+Bt)\sin(\omega t+\theta)\right]\cdot\sin(\omega t+\theta)\ ,\\
\frac{\partial S}{\partial B} &= -2\cdot\sum_{t=1}^{n}\left[z_t-(A+Bt)\sin(\omega t+\theta)\right]\cdot t\sin(\omega t+\theta)\ ,\\
\frac{\partial S}{\partial \theta} &= -2\cdot\sum_{t=1}^{n}\left[z_t-(A+Bt)\sin(\omega t+\theta)\right]\cdot(A+Bt)\cos(\omega t+\theta)\ ,\\
\frac{\partial S}{\partial \omega} &= -2\cdot\sum_{t=1}^{n}\left[z_t-(A+Bt)\sin(\omega t+\theta)\right]\cdot(At+Bt^2)\cos(\omega t+\theta)\ .
\end{aligned} \tag{4}$$

Korzystając z warunku koniecznego istnienia ekstremym funkcji F, po przekształceniach otrzymujemy układ równań normalnych dla modelu (2):

$$\begin{cases}
\hat{A}\cdot\sum_{t=1}^{n}\sin^2\left(\hat{\omega}t+\hat{\theta}\right)+\hat{B}\cdot\sum_{t=1}^{n}t\sin^2\left(\hat{\omega}t+\hat{\theta}\right)=\sum_{t=1}^{n}z_t\sin\left(\hat{\omega}t+\hat{\theta}\right)\ ;\\
\hat{A}\cdot\sum_{t=1}^{n}t\sin^2\left(\hat{\omega}t+\hat{\theta}\right)+\hat{B}\cdot\sum_{t=1}^{n}t^2\sin^2\left(\hat{\omega}t+\hat{\theta}\right)=\sum_{t=1}^{n}t\,z_t\sin\left(\hat{\omega}t+\hat{\theta}\right)\ ;\\
\frac{1}{2}\hat{A}^2\cdot\sum_{t=1}^{n}\sin\left(2\hat{\omega}t+2\hat{\theta}\right)+\hat{A}\hat{B}\cdot\sum_{t=1}^{n}t\sin\left(2\hat{\omega}t+2\hat{\theta}\right)+\frac{1}{2}\hat{B}^2\cdot\sum_{t=1}^{n}t^2\sin\left(2\hat{\omega}t+2\hat{\theta}\right)=\\
\qquad=\hat{A}\cdot\sum_{t=1}^{n}z_t\cos\left(\hat{\omega}t+\hat{\theta}\right)+\hat{B}\cdot\sum_{t=1}^{n}t\,z_t\cos\left(\hat{\omega}t+\hat{\theta}\right)\ ;\\
\frac{1}{2}\hat{A}^2\cdot\sum_{t=1}^{n}t\sin\left(2\hat{\omega}t+2\hat{\theta}\right)+\hat{A}\hat{B}\cdot\sum_{t=1}^{n}t^2\sin\left(2\hat{\omega}t+2\hat{\theta}\right)+\frac{1}{2}\hat{B}^2\cdot\sum_{t=1}^{n}t^3\sin\left(2\hat{\omega}t+2\hat{\theta}\right)=\\
\qquad=\hat{A}\cdot\sum_{t=1}^{n}t\,z_t\cos\left(\hat{\omega}t+\hat{\theta}\right)+\hat{B}\cdot\sum_{t=1}^{n}t^2z_t\cos\left(\hat{\omega}t+\hat{\theta}\right)\ .
\end{cases} \tag{5}$$

Układ równań (5) ma skomplikowaną postać. Rozwiązanie tego układu, czyli obliczenie wartości ocen parametrów A, B, ω, θ, jest zadaniem bardzo trudnym. Przezwyciężenie tej trudności pozwoliłoby na wyznaczenie okresu wahań T na podstawie danych empirycznych. Jest to odrębne i niełatwe zagadnienie. W dalszym ciągu niniejszych rozważań nałożymy pewne ograniczenia na dane empiryczne. Ograniczenia te pozwolą znacznie uprościć układ równań (5) i dokonać estymacji szukanych parametrów. Będziemy od tego momentu za-

kładać, że znany jest okres $\hat{T}$ (czyli także częstość $\hat{\omega}$) badanego zjawiska. Ma to miejsce przy opisie sezonowości, wtedy dla danych miesięcznych $\hat{T} = 12$, a dla kwartalnych $\hat{T} = 4$. Przyjmujemy także dwa następujące założenia: 1° dysponujemy kolejnymi punktami empirycznymi (z_t, t) $(t = 1, 2, ..., n)$, znaczy to, że szereg czasowy jest kompletny, nie ma brakujących danych empirycznych; 2° liczba punktów empirycznych jest całkowitą wielokrotnością okresu badanego zjawiska, tzn. $n = k\hat{T}$ – łatwo to zapewnić przez pominięcie w obliczeniach co najwyżej kilku punktów empirycznych.

Przy obecnej sprawozdawczości i szerokim dostępie do danych można uznać, że powyższe restrykcje nie powodują istotnych ograniczeń zastosowania proponowanych w tym artykule rozwiązań.

Dalsze rozważania będą miały na celu uproszczenie układu równań (5). Ze względu na to, że założyliśmy znajomość okresu $\hat{T}$, czwarte równanie w układzie (5) należy pominąć. Sumy występujące po lewych stronach pozostałych równań obliczamy wykorzystując odpowiednie wzory indukcyjne. W czterech przypadkach zostało to już pokazane w pracy [7], przytoczymy tutaj ostateczne wzory:

$$S_1 = \sum_{t=1}^{n} \sin(2\hat{\omega}t + 2\hat{\theta}) = \frac{\sin(n\hat{\omega})\cdot\sin[(n+1)\hat{\omega}+2\hat{\theta}]}{\sin\hat{\omega}},$$

$$S_2 = \sum_{t=1}^{n} \sin^2\left(\hat{\omega}t + \hat{\theta}\right) = \frac{n}{2} - \frac{\sin(n\hat{\omega})\cdot\cos[(n+1)\hat{\omega}+2\hat{\theta}]}{2\sin\hat{\omega}},$$

$$S_3 = \sum_{t=1}^{n} t\cdot\sin(2\hat{\omega}t + 2\hat{\theta}) = \frac{(n+1)\sin(2n\hat{\omega}+2\hat{\theta}) - n\sin[2(n+1)\hat{\omega}+2\hat{\theta}] - \sin 2\hat{\theta}}{4\sin^2\hat{\omega}},$$

$$S_4 = \sum_{t=1}^{n} t\cdot\sin^2(\hat{\omega}t + \hat{\theta}) = \frac{n(n+1)}{4} - \frac{(n+1)\cos(2n\hat{\omega}+2\hat{\theta}) - n\cos[2(n+1)\hat{\omega}+2\hat{\theta}] - \cos 2\hat{\theta}}{8\sin^2\hat{\omega}}.$$

Obliczenie kolejnych sum wymaga zastosowania wzoru, którego prawdziwości potrafimy dowieść. Wzór ten ma postać:

$$\sum_{k=1}^{n} k^2 a^k = \frac{n^2 a^{n+3} - (2n^2+2n-1)a^{n+2} + (n^2+2n+1)a^{n+1} - a^2 - a}{(a-1)^3}, \quad \text{dla } n \geq 1 \text{ i } a \neq 1.$$

Po podstawieniu do niego $a=\cos x+i\sin x$ i dokonaniu odpowiednich przekształceń, otrzymujemy dwa nowe i bardzo praktyczne wzory indukcyjne:

$$\sum_{k=1}^{n} k^2\cos kx=\frac{-n^2\sin\left(\frac{2n+3}{2}x\right)+(2n^2+2n-1)\sin\left(\frac{2n+1}{2}x\right)-(n^2+2n+1)\sin\left(\frac{2n-1}{2}x\right)}{8\sin^3\frac{x}{2}},$$

$$\sum_{k=1}^{n} k^2\sin kx=\frac{n^2\cos\left(\frac{2n+3}{2}x\right)-(2n^2+2n-1)\cos\left(\frac{2n+1}{2}x\right)+(n^2+2n+1)\cos\left(\frac{2n-1}{2}x\right)-2\cos\frac{x}{2}}{8\sin^3\frac{x}{2}}.$$

Wykorzystując te związki obliczamy dwie ostatnie sumy brakujące w układzie równań (5). Mamy więc:

$$S_5=\sum_{t=1}^{n} t^2\sin(2\hat{\omega}t+2\hat{\theta})=$$

$$=\frac{n^2\cos[(2n+3)\hat{\omega}+2\hat{\theta}]-(2n^2+2n-1)\cos[(2n+1)\hat{\omega}+2\hat{\theta}]+(n^2+2n+1)\cos[(2n-1)\hat{\omega}+2\hat{\theta}]-2\cos\hat{\omega}\cos 2\hat{\theta}}{8\sin^3\hat{\omega}},$$

$$S_6=\sum_{t=1}^{n} t^2\sin^2(\hat{\omega}t+\hat{\theta})=\frac{n(n+1)(2n+1)}{12}+$$

$$+\frac{n^2\sin[(2n+3)\hat{\omega}+2\hat{\theta}]-(2n^2+2n-1)\sin[(2n+1)\hat{\omega}+2\hat{\theta}]+(n^2+2n+1)\sin[(2n-1)\hat{\omega}+2\hat{\theta}]-2\cos\hat{\omega}\sin 2\hat{\theta}}{16\sin^3\hat{\omega}}.$$

Jeśli uwzględnimy poczynione wcześniej założenie $n=k\hat{T}$, powyższe wzory przybiorą znacznie prostszą postać. Ponieważ $\hat{\omega}=2\pi/\hat{T}$, więc $n\hat{\omega}=2k\pi$, wobec tego po dokonaniu elementarnych przekształceń otrzymujemy:

$$S_1=0,\quad S_2=\frac{n}{2},\quad S_3=\frac{-n\cos(\hat{\omega}+2\hat{\theta})}{2\sin\hat{\omega}},\quad S_4=\frac{n(n+1)}{4}-\frac{n\sin(\hat{\omega}+2\hat{\theta})}{4\sin\hat{\omega}},$$

$$S_5=\frac{(n^2+n)\sin 2\hat{\theta}-n^2\cos\hat{\omega}\sin(\hat{\omega}+2\hat{\theta})}{2\sin^2\hat{\omega}},\quad S_6=\frac{n(n+1)(2n+1)}{12}+\frac{n^2\cos(2\hat{\omega}+2\hat{\theta})-(n^2+2n)\cos 2\hat{\theta}}{8\sin^2\hat{\omega}}.$$

Uwzględniając powyższe związki i pomijając czwarte równanie, układ równań normalnych zapisujemy w postaci:

$$\begin{cases} \hat{A}\cdot\frac{n}{2}+\hat{B}\cdot\left[\frac{n(n+1)}{4}-\frac{n\sin(\hat{\omega}+2\hat{\theta})}{4\sin\hat{\omega}}\right]=\sum_{t=1}^{n}z_t\sin\left(\hat{\omega}t+\hat{\theta}\right); \\ \hat{A}\cdot\left[\frac{n(n+1)}{4}-\frac{n\sin(\hat{\omega}+2\hat{\theta})}{4\sin\hat{\omega}}\right]+\hat{B}\cdot\left[\frac{n(n+1)(2n+1)}{12}+\frac{n^2\cos(2\hat{\omega}+2\hat{\theta})-(n^2+2n)\cos 2\hat{\theta}}{8\sin^2\hat{\omega}}\right]=\sum_{t=1}^{n}t\,z_t\sin\left(\hat{\omega}t+\hat{\theta}\right); \\ \hat{A}\hat{B}\cdot\frac{-n\cos(\hat{\omega}+2\hat{\theta})}{2\sin\hat{\omega}}+\hat{B}^2\cdot\frac{(n^2+n)\sin 2\hat{\theta}-n^2\cos\hat{\omega}\sin(\hat{\omega}+2\hat{\theta})}{4\sin^2\hat{\omega}}=\hat{A}\cdot\sum_{t=1}^{n}z_t\cos\left(\hat{\omega}t+\hat{\theta}\right)+\hat{B}\cdot\sum_{t=1}^{n}t\,z_t\cos\left(\hat{\omega}t+\hat{\theta}\right). \end{cases} \tag{6}$$

W układzie (6) ω jest znanym parametrem, a niewiadomymi są: A, B, θ. Układ ten, mimo swojej skomplikowanej postaci, daje możliwość estymacji parametrów A, B, θ. Można go rozwiązać wykorzystując odpowiedni program komputerowy, pozwalający na numeryczne rozwiązywanie układów równań.

Estymacja składnika wahań sezonowych

Układ równań (6) jest szczególnie przydatny w przypadku opisu zmienności sezonowej, ponieważ znamy wtedy okres wahań $\hat{T}$. W przypadku obserwacji miesięcznych $\hat{T}=12$, czyli $\hat{\omega}=2\pi/\hat{T}=\pi/6$, a w przypadku obserwacji kwartalnych $\hat{T}=4$, więc $\hat{\omega}=2\pi/\hat{T}=\pi/2$. Gdy dysponujemy obserwacjami miesięcznymi, po podstawieniu $\hat{\omega}=\pi/6$ i niezbędnych uproszczeniach, układ równań (6) przybiera postać:

$$\begin{cases} \hat{A}\cdot\frac{n}{2}+\hat{B}\cdot\left[\frac{n(n+1)}{4}-\frac{n}{4}\cos 2\hat{\theta}-\frac{n\sqrt{3}}{4}\sin 2\hat{\theta}\right]=\sum_{t=1}^{n}z_t\sin\left(\frac{\pi}{6}t+\hat{\theta}\right); \\ \hat{A}\cdot\left[\frac{n(n+1)}{4}-\frac{n}{4}\cos 2\hat{\theta}-\frac{n\sqrt{3}}{4}\sin 2\hat{\theta}\right]+\hat{B}\cdot\left[\frac{n(n+1)(2n+1)}{12}-\frac{n(n+4)}{4}\cos 2\hat{\theta}-\frac{n^2\sqrt{3}}{4}\sin 2\hat{\theta}\right]= \\ \qquad=\sum_{t=1}^{n}t\,z_t\sin\left(\frac{\pi}{6}t+\hat{\theta}\right); \\ \hat{A}\hat{B}\cdot\left[\frac{n}{2}\sin 2\hat{\theta}-\frac{n\sqrt{3}}{2}\cos 2\hat{\theta}\right]+\hat{B}^2\cdot\left[\frac{n(n+4)}{4}\sin 2\hat{\theta}-\frac{n^2\sqrt{3}}{4}\cos 2\hat{\theta}\right]= \\ \qquad=\hat{A}\cdot\sum_{t=1}^{n}z_t\cos\left(\frac{\pi}{6}t+\hat{\theta}\right)+\hat{B}\cdot\sum_{t=1}^{n}t\,z_t\cos\left(\frac{\pi}{6}t+\hat{\theta}\right). \end{cases} \tag{7}$$

Rozważany i oszacowany model (2) z kolei wykorzystamy do opisu i prognozy skupu mleka w Polsce. Jednocześnie, wykorzystując te same dane empiryczne, oszacujemy parametry modelu o stałej amplitudzie:

$$z_t = s + A \cdot \sin(\omega\, t + \theta) + \varepsilon_t \,. \tag{8}$$

Estymację parametrów modelu (8) przeprowadzono w pracy [7]. Gdy okres $\hat{T}$ opisywanego zjawiska jest znany i $n = k\hat{T}$, wtedy oszacowanie parametrów modelu (8) jest bardzo proste. Mamy mianowicie:

$$\hat{s} = \frac{1}{n} \cdot \sum_{t=1}^{n} z_t; \qquad \hat{\theta} = \operatorname{arctg}\left[\sum_{t=1}^{n} z_t \cos \hat{\omega}\, t \Big/ \sum_{t=1}^{n} z_t \sin \hat{\omega}\, t \right];$$

$$\hat{A} = \frac{2}{n} \cdot \left[\cos\hat{\theta} \cdot \sum_{t=1}^{n} z_t \sin \hat{\omega}\, t + \sin\hat{\theta} \cdot \sum_{t=1}^{n} z_t \cos \hat{\omega}\, t \right]. \tag{9}$$

Zastosowanie praktyczne proponowanego modelu

W tabeli 1 podano dane dotyczące skupu mleka w Polsce w latach 1969–1979. Są to dane miesięczne (w mln litrów), które zaczerpnięto z pracy [7, s. 455] oraz z Biuletynów Statystycznych z różnych lat. Korzystamy z tych danych, gdyż wahania sezonowe skupu mleka w latach 1969–1979 charakteryzowały się właśnie rosnącą amplitudą. Ponadto, chcemy dokonać porównania z wynikami otrzymanymi w pracy [7].

Do estymacji parametrów modelu (2) wykorzystujemy dane dotyczące skupu mleka w Polsce z lat 1970–1977, pozostałe dane posłużą do oceny wyznaczonych prognoz. Trend opiszemy funcją liniową $f(t) = a + bt$. Jej parametry szacujemy metodą średnich. Odpowiedni układ równań ma postać:

$$1176 \cdot \hat{b} + 48 \cdot \hat{a} = 24734{,}9 \,,$$

$$3480 \cdot \hat{b} + 48 \cdot \hat{a} = 34680{,}1 \,.$$

Po jego rozwiązaniu i zaokrągleniach otrzymujemy:

$$\hat{f}(t) = 409{,}5563 + 4{,}3165 \cdot t, \quad \text{gdzie } t = 1 \equiv \text{styczeń 1970 r.}$$

Następnie obliczamy różnice $z_t = y_t - \hat{f}(t)$ dla $t = 1 \equiv$ I 1970 r. do $t = 96 \equiv$ XII 1977 r. i za ich pomocą obliczamy potrzebne sumy:

Tabela 1
Skup mleka w Polsce w latach 1969–1979 (w mln l)

Miesiąc \ Rok	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979
I	346,8	336,5	353,1	410,7	476,6	564,6	514,9	588,9	641,5	656,8	618,0
II	328,1	319,6	339,1	407,8	450,4	527,5	481,3	561,8	604,4	605,3	557,8
III	376,9	369,8	381,4	457,5	522,1	596,8	555,5	633,0	704,7	713,6	680,4
IV	380,8	391,3	392,7	473,9	520,3	594,1	581,2	635,1	715,9	741,6	713,0
V	470,6	475,1	522,4	637,6	699,6	752,3	783,3	806,4	933,4	953,2	895,2
VI	590,1	596,4	612,2	726,1	806,5	911,1	907,3	983,8	1043,7	1113,0	1055,0
VII	573,4	591,5	597,8	682,6	777,1	888,9	879,6	946,7	1073,5	1123,0	1122,7
VIII	481	559,4	525,0	654,6	744,0	807,9	828,3	921,1	994,5	1043,9	1073,1
IX	444,2	523,4	507,2	638,7	710,4	759,1	784,4	827,6	887,1	922,3	980,3
X	419,5	463,5	470,9	599,5	638,9	660,7	702,8	720,5	818,2	814,9	912,0
XI	326	351,1	375,1	469,2	514,3	515,2	541,0	612,7	642,1	674,8	691,1
XII	303,4	329,4	379,4	439,2	514,0	484,8	550,1	594,4	616,4	637,6	679,2
Skup roczny	5040,8	5307,0	5456,3	6597,4	7374,2	8063,0	8109,7	8832,0	9675,4	10000,0	9977,8

Źródło: [7, s. 455] i opracowanie własne na podstawie Biuletynów Statystycznych z różnych lat.

$$\sum_{t=1}^{96} z_t \sin\left(\frac{\pi}{6}t\right) = -3820{,}8837\,; \qquad \sum_{t=1}^{96} z_t \cos\left(\frac{\pi}{6}t\right) = -7258{,}3966\,;$$

$$\sum_{t=1}^{96} t\, z_t \sin\left(\frac{\pi}{6}t\right) = -194246{,}6642\,; \qquad \sum_{t=1}^{96} t\, z_t \cos\left(\frac{\pi}{6}t\right) = -395983{,}5646\,.$$

Po podstawieniu ich do (7) układ równań normalnych w naszym przykładzie przybiera następującą postać:

$$\begin{cases} 48\cdot\hat{A} + (2328 - 24\cdot\cos 2\hat{\theta} - 41{,}5692\cdot\sin 2\hat{\theta})\cdot\hat{B} = -3820{,}8837\cdot\cos\hat{\theta} - 7258{,}3966\cdot\sin\hat{\theta}; & (10) \\ (2328 - 24\cdot\cos 2\hat{\theta} - 41{,}5692\cdot\sin 2\hat{\theta})\cdot\hat{A} + \\ + (149768 - 2400\cdot\cos 2\hat{\theta} - 3990{,}6451\cdot\sin 2\hat{\theta})\cdot\hat{B} = -194246{,}6642\cdot\cos\hat{\theta} - 395983{,}5646\cdot\sin\hat{\theta}; \\ (48\cdot\sin 2\hat{\theta} - 83{,}1384\cdot\cos 2\hat{\theta})\cdot\hat{A}\cdot\hat{B} + (2400\sin 2\hat{\theta} - 3990{,}6451\cdot\cos 2\hat{\theta})\cdot\hat{B}^2 = \\ = -7258{,}3966\cdot\hat{A}\cdot\cos\hat{\theta} + 3820{,}8837\cdot\hat{A}\cdot\sin\hat{\theta} - 395983{,}5646\cdot\hat{B}\cdot\cos\hat{\theta} + 194246{,}6642\cdot\hat{B}\cdot\sin\hat{\theta}. \end{cases}$$

Poszukujemy jego rozwiązania dla $\hat{\theta} \in [-\pi;\ \pi]$.

Układ równań (10) rozwiązano przy użyciu programu komputerowego *Mathematica 3.0*. Z rozwiązań równoważnych wybrano:

$$\hat{A} = 110{,}24, \quad \hat{B} = 1{,}26178, \quad \hat{\theta} = -2{,}05727.$$

Ostatecznie, po zaokrągleniach, oszacowany model dla skupu mleka w Polsce przyjmuje postać:

$$\hat{y}(t) = 409{,}57 + 4{,}317\cdot t + (110{,}24 + 1{,}262\cdot t)\cdot\sin\left(\frac{\pi}{6}t - 2{,}057\right), \qquad (11)$$

w którym $t = 1 \equiv$ styczeń 1970 r.

Miary dopasowania modelu (11) w przedziale jego aproksymacji kształtują się następująco:

współczynnik determinacji $R^2 = 1 - \left[\sum_{t=1}^{96}(y_t - \hat{y}_t) \Big/ \sum_{t=1}^{96}(y_t - \bar{y})\right] = 0{,}9404\,,$

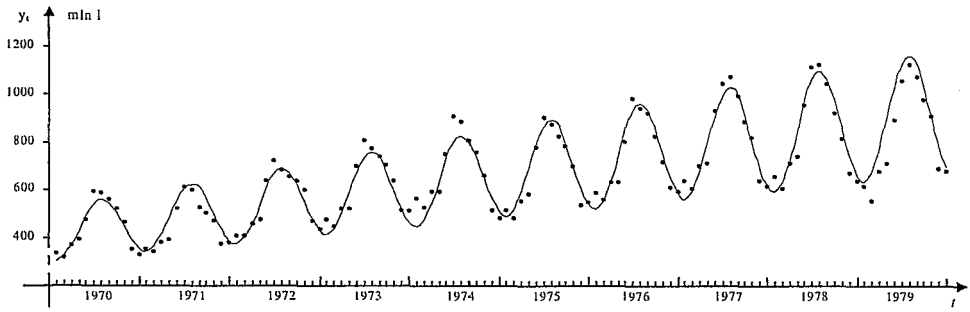
który interpretujemy w ten sposób, że oszacowany model (11) objaśnia w 94% zmienność badanej zmiennej zależnej; odchylenie standardowe składnika resztowego $s_e = \sqrt{\sum_{t=1}^{96}(y_t - \hat{y}_t)^2 \Big/ (96-5)} = 44{,}8$ mln l; współczynnik zmienności przypadkowej $v = 100 s_e / \bar{y} = 7{,}3\%$.

Na rysunku 1 zamieszczono dane empiryczne i wykres oszacowanego modelu (11).

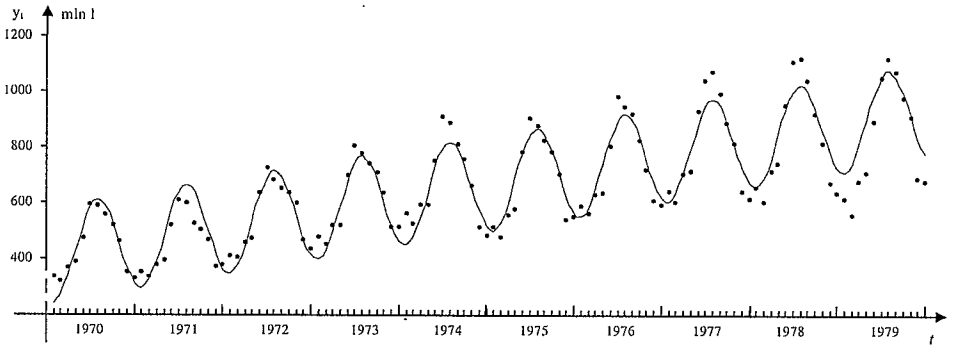
Dla porównania oszacujemy jeszcze model (8) o stałej amplitudzie. Funkcja trendu jest taka sama jak poprzednio, a parametry składnika wahań sezonowych szacujemy za pomocą wzorów (9). Otrzymujemy: $\hat{s} = 16 \cdot 10^{-9}$, $\hat{\theta} = 1{,}086246$, $\hat{A} = -170{,}88855$. Przekonujemy się więc, że estymacja parametrów trendu metodą średnich powoduje zmniejszenie o jeden liczby parametrów harmoniki. Wtedy we wzorach (9) zawsze otrzymujemy $\hat{s} = 0$. Ostatecznie, po zaokrągleniach, model ze stałą amplitudą przyjmuje postać:

$$\hat{y}(t) = 409{,}56 + 4{,}317 \cdot t + 170{,}889 \cdot \sin\left(\frac{\pi}{6} t - 2{,}056\right), \qquad (12)$$

w którym $t = 1 \equiv$ styczeń 1970 r. Miary dopasowania modelu (12) do tych samych danych empirycznych są następujące: $R^2 = 0{,}921$, $s_e = 51{,}5$, $v = 8{,}3\%$, a jego wykres zamieszczono na rysunku 2.



Rysunek 1
Skup mleka w Polsce w latach 1970–1979 (model o rosnącej amplitudzie, prognoza na lata 1978 i 1979)



Rysunek 2
Skup mleka w Polsce w latach 1970–1979 (model o stałej amplitudzie, prognoza na lata 1978 i 1979)

Prognozy obliczone na podstawie wzorów (11) i (12) na rok 1978 oraz kres górny błędu względnego wyrażony w procentach dla każdego miesiąca podane są w tabelach 2 oraz 3. W tabeli 4 pokazana została także prognoza wsteczna na rok 1969 obliczona na podstawie (11).

Tabela 2
Prognoza na rok 1978 otrzymana za pomocą modelu o rosnącej amplitudzie (11)

1978	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
y_t	656,8	605,3	713,6	741,6	953,2	1113,0	1123,0	1043,9	922,3	814,9	674,8	637,6
$\hat{y}_t$	595,8	634,5	727,0	850,0	971,9	1061,1	1094,2	1063,0	976,3	858,0	741,0	657,8
Kgbw%	**9,29**	**4,82**	**1,87**	**14,61**	**1,96**	**4,67**	**2,57**	**1,83**	**5,85**	**5,29**	**9,81**	**3,17**

Tabela 3
Prognoza na rok 1978 otrzymana za pomocą modelu o stałej amplitudzie (12)

1978	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
y_t	656,8	605,3	713,6	741,6	953,2	1113,0	1123,0	1043,9	922,3	814,9	674,8	637,6
$\hat{y}_t$	657,5	688,0	757,3	847,9	936,7	1001,0	1024,9	1003,0	942,4	860,4	780,3	724,5
Kgbw%	**0,11**	**13,67**	**6,12**	**14,33**	**1,73**	**10,06**	**8,73**	**3,92**	**2,18**	**5,59**	**15,63**	**13,63**

Tabela 4
Prognoza wsteczna na rok 1969 otrzymana za pomocą modelu o rosnącej amplitudzie (11)

1969	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
y_t	346,8	328,1	376,9	380,8	470,6	590,1	573,4	481,0	444,2	419,5	326,0	303,4
$\hat{y}_t$	265,8	283,7	324,5	378,7	433,3	474,4	491,8	481,4	446,4	396,9	347,3	312,1
Kgbw%	**23,36**	**13,53**	**13,91**	**0,54**	**7,93**	**19,60**	**14,23**	**0,08**	**0,49**	**5,38**	**6,53**	**2,87**

Podsumowanie

Model (11) o rosnącej monotonicznie amplitudzie sezonowości jest odpowiednim modelem do opisu skupu mleka w Polsce, odzwierciedla zasadnicze cechy tej zmiennej zależnej, ale nie eliminuje oczywistych w tej sytuacji wahań przypadkowych. Stanowi pewien postęp w stosunku do przytoczonego modelu (12) o stałej amplitudzie sezonowości. Zapewnia zadowalającą prognozę długoterminową. Model (12) ma bardzo prostą estymację parametrów (9) i był wielokrotnie wykorzystywany do opisu odpowiadających mu teoretycznie zjawisk klimatycznych. Estymacja parametrów modelu (2) ze zmienną amplitudą jest zadaniem złożonym i pracochłonnym, ale wyprowadzenie odpowiadających mu równań normalnych (6) i ich uproszczeń przyszłego użytkownika nie musi interesować, może skorzystać z ostatecznych propozycji (7). Pamiętamy też o tym, że ten sam problem można rozwiązywać w różny sposób, ale zawsze należy stosować metody zgodne z surowymi regułami postępowania statystycznego.

Literatura

BOX G.E.P., JENKINS G.M., 1983: *Analiza szeregów czasowych. Prognoza i sterowanie*. PWN, Warszawa. [1]

CHOW G.C., 1995: *Ekonometria*. PWN, Warszawa. [2]

CIEŚLAK M. (red. naukowa), 1997: *Prognozowanie gospodarcze. Metody i zastosowania*. PWN, Warszawa. [3]

MILO W., 1990: *Nieliniowe modele ekonometryczne*. PWN, Warszawa. [4]

RADZIKOWSKA B. (red.), 2000: *Metody prognozowania. Zbiór zadań*. Wydawnictwo AE we Wrocławiu, Wrocław. [5]

RYŻYK I.M., GRADSZTEJN I.S., 1964: *Tablice całek, sum, szeregów i iloczynów*. PWN, Warszawa. [6]
SMOLIK S., 1995: *Uproszczona procedura estymacjimodelu wahań okresowych*. „Przegląd Statystyczny", R. XLII, z. 3–4. [7]
SMOLIK S., 1998: *Oszczędne modele dla okresowych szeregów czasowych*. [w:] Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych. Red. A. Zeliaś, Akademia Ekonomiczna w Krakowie, s.177–188. [8]
WELFE W., WELFE A., 1996: *Ekonometria stosowana*. PWN, Warszawa. [9]
ZAJĄC K., 1994: *Zarys metod statystycznych*, PWE, Warszawa. [10]
ZELIAŚ A., 1997: *Teoria prognozy*. Wyd. III zmienione, PWE, Warszawa. [11]

Description of periodic phenomena with a monotonically changing amplitude

Abstract

For the purposes of description of economic and natural phenomena with a distinctly outlined trend and a dominating seasonal dependence, the model was proposed: $y_t = f(t) + (A + B \cdot t) \cdot \sin(\omega\ t + \theta) - \varepsilon_t \quad (t = 1, 2, ..., n)$. The analytic shape of the trend function $f(t)$, in its general form, is adjusted to the phenomenon being described and subsequently, its parameters are estimated with the average method, using empirical data (t, y_t). In such a case, seasonal relations with changing amplitude can sufficiently be described with the model:

$$z_t = (A + B \cdot t) \cdot \sin(\omega\ t + \theta) - \varepsilon_t, \text{ whereas } z_t = y_t - \hat{f}(t) \ (t = 1, 2, ..., n)$$

and $\omega = 2\pi / T$. Whilst using monthly observations $\hat{T} = 12$ and $n = k\hat{T}$, this model's parameters are liable to estimation out of the following equation system:

$$\begin{cases} \hat{A}\cdot\frac{n}{2}+\hat{B}\cdot\left[\frac{n(n+1)}{4}-\frac{n}{4}\cos 2\hat{\theta}-\frac{n\sqrt{3}}{4}\sin 2\hat{\theta}\right]=\sum_{t=1}^{n} z_t \sin\left(\frac{\pi}{6}t+\hat{\theta}\right); \\ \hat{A}\cdot\left[\frac{n(n+1)}{4}-\frac{n}{4}\cos 2\hat{\theta}-\frac{n\sqrt{3}}{4}\sin 2\hat{\theta}\right]+\hat{B}\cdot\left[\frac{n(n+1)(2n+1)}{12}-\frac{n(n+4)}{4}\cos 2\hat{\theta}-\frac{n^2\sqrt{3}}{4}\sin 2\hat{\theta}\right]= \\ \qquad =\sum_{t=1}^{n} t\, z_t \sin\left(\frac{\pi}{6}t+\hat{\theta}\right); \\ \hat{A}\hat{B}\cdot\left[\frac{n}{2}\sin 2\hat{\theta}-\frac{n\sqrt{3}}{2}\cos 2\hat{\theta}\right]+\hat{B}^2\cdot\left[\frac{n(n+4)}{4}\sin 2\hat{\theta}-\frac{n^2\sqrt{3}}{4}\cos 2\hat{\theta}\right]= \\ \qquad =\hat{A}\cdot\sum_{t=1}^{n} z_t \cos\left(\frac{\pi}{6}t+\hat{\theta}\right)+\hat{B}\cdot\sum_{t=1}^{n} t\, z_t \cos\left(\frac{\pi}{6}t+\hat{\theta}\right). \end{cases}$$

This model has been employed, with a linear tendency, in our description and forecast of milk purchase in Poland.

Sylwester Smolik
Katedra Zastosowań Matematyki SGGW

Modelowanie statystyczne zjawisk z trendem i sezonowością

Wprowadzenie

Ekonometrycy analizując szeregi czasowe zjawisk ekonomicznych wyróżniają w nich następujące składowe: trend, wahania sezonowe, wahania koniunkturalne oraz wahania losowe. Wszystkie z wymienionych wahań nie muszą jednocześnie występować w opisywanym procesie. Przez trend rozumie się pewną wypośrodkowującą linię wyznaczającą zasadniczy kierunek rozwojowy badanego zjawiska. Nie wyjaśnia on, dlaczego badana wielkość ulegała zmianom, jednak posługiwanie się trendem do opisu przeszłości lub przewidywania przyszłości jest rozpowszechnione i celowe. Wahania sezonowe obrazują zmiany zjawiska w okresie rocznym. Mogą być stałe, tzn. systematycznie z roku na rok w odpowiadających okresach zakłócają trend w ten sam sposób. Wahania sezonowe mogą też być zmienne, tzn. w każdym następnym roku rośnie (lub maleje) ich nasilenie – zachowując charakter zakłóceń. Najczęściej objawia się to zjawisko rosnącą amplitudą zakłóceń wraz ze wzrostem wartości trendu. Sprawcą sezonowości zjawisk ekonomicznych są czynniki klimatyczno-przyrodnicze oraz powtarzające się systematycznie zdarzenia kalendarzowe (święta). Do najważniejszych z nich należą: czas trwania dnia, temperatura powietrza, opady, oraz zjawiska wtórne w stosunku do wymienionych (grubość pokrywy śnieżnej lub grubość pokrywy lodowej na rzekach). Wszystkie tzw. podstawowe czynniki sezonowości można wyrazić ilościowo za pomocą liczb i charakteryzują się one następującymi własnościami:
1° są funkcjami okresowymi czasu o okresie rocznym;
2° ich wartości, jak i kształt są zróżnicowane przede wszystkim w zależności od położenia geograficznego.

Podstawowe czynniki kształtujące sezonowość nie muszą oddziaływać na rozpatrywane zjawisko ekonomiczne w sposób bezpośredni, mogą występować przesunięcia czasowe i łańcuchy powiązań wieloogniwowych.

Wahania koniunkturalne nie muszą być ściśle okresowymi, trwają od kilku do kilkunastu lat. Zwykle wahania dotyczące rozwoju społeczno-gospodarczego krajów mają dłuższy okres, podczas gdy okresy dotyczące poszczególnych zjawisk są zwykle krótsze (np. cykl świński). Wahania losowe traktujemy jako wynik działania dużej liczby przyczyn ubocznych, mających charakter podobny do błędów przypadkowych, których analizą zajmuje się statystyka matematyczna. Zdarza się także, że w szeregu czasowym daje się wyodrębnić wahania katastrofalne. Występują one sporadycznie i powodują znacznie większe odchylenia od trendu w porównaniu z wahaniami losowymi. Są one spowodowane przez zdarzenia historyczne, takie jak wojna, kryzys, katastrofy żywiołowe, epidemie itp. Mając dany szereg czasowy, chcemy racjonalnie rozłożyć go na opisane wyżej składowe oraz poprawnie oszacować parametry opisującego go modelu.

Estymacja parametrów proponowanego modelu

Z wahaniami sezonowymi spotykamy się bardzo często w rolnictwie, które jest uzależnione od przebiegu pór roku. Zmianom tym podlega zapotrzebowanie na siłę roboczą, siłę pociągową, skup mleka, żywca, płodów rolnych i związany z nimi transport kolejowy towarów. Na potrzeby rolnictwa prowadzi się systematyczne obserwacje meteorologiczne. Ich efektem są publikowane średnie miesięczne temperatury powietrza i gruntu, opady, usłonecznienie itp. wybranych miejscowości. W powszechnie dostępnych periodykach są publikowane dane dotyczące miesięcznej produkcji najważniejszych towarów i różnych parametrów opisujących stan gospodarki. Opracujemy model opisujący tego typu zjawiska. Zakładamy, że wartości opisywanej zmiennej zależnej Y_{it} charakteryzują się składową i-tego roku i mają wyraźną sezonowość zależną od t-tego miesiąca, ponadto są obarczone błędem losowym. Mamy zatem:

$$Y_{it} = L(i) + M(t) + \varepsilon_{it} \qquad (i = 1,2,..., l;\ t = 1,2,..., 12). \tag{1}$$

Bez względu na postać funkcji L, zawsze $L(i) = L_i$ dla każdego i. Natomiast zakładamy, że sezonowość daje się opisać jedną harmoniką. Zauważmy, że w naszym przypadku pulsacja harmoniki $w = 2\pi/T = 2\pi/12 = \pi/6$, wtedy proponowany model przyjmuje postać:

$$y_{it} = B_i + A \cdot \sin\left(\frac{\pi}{6}t + \theta\right) + \varepsilon_{it}, \tag{2}$$

w którym A nazywamy amplitudą, a θ fazą początkową harmoniki i razem z B_i ($i = 1, 2, ..., l$) stanowią szacowane parametry modelu (2), na podstawie wyników obserwacji zebranych w macierz liczb o l wierszach i $t = 12$ kolumnach. Zakładamy a priori, że ε_{it} są niezależnymi zmiennymi losowymi o rozkładzie normalnym z wartością oczekiwaną zero i wariancją σ^2 dla wszystkich i, t. Oszacujemy parametry modelu (2) metodą najmniejszych kwadratów (MNK). Mamy więc:

$$S(B_i, \theta, A) = \sum_{i=1}^{l} \sum_{t=1}^{12} \varepsilon_{it}^2 = \sum_{i=1}^{l} \sum_{t=1}^{12} \left[y_{it} - B_i - A\sin\left(\frac{\pi}{6}t + \theta\right) \right]^2 = \min. \qquad (3)$$

Korzystamy z warunku koniecznego ekstremum funkcji $S(B_i, \theta, A)$:

$$\frac{\partial S}{\partial B_i} = 2 \cdot \sum_{t=1}^{12} \left[y_{it} - B_i - A\sin\left(\frac{\pi}{6}t + \theta\right) \right] \cdot (-1) = 0, \qquad \text{dla } i = 1, 2, ..., l;$$

$$\frac{\partial S}{\partial \theta} = 2 \cdot \sum_{i=1}^{l} \sum_{t=1}^{12} \left[y_{it} - B_i - A\sin\left(\frac{\pi}{6}t + \theta\right) \right] \cdot \left[-A\cos\left(\frac{\pi}{6}t + \theta\right) \right] = 0; \qquad (4)$$

$$\frac{\partial S}{\partial A} = 2 \cdot \sum_{i=1}^{l} \sum_{t=1}^{12} \left[y_{it} - B_i - A\sin\left(\frac{\pi}{6}t + \theta\right) \right] \cdot \left[-\sin\left(\frac{\pi}{6}t + \theta\right) \right] = 0.$$

Po prostych przekształceniach otrzymujemy układ równań normalnych dla modelu (2):

$$\begin{cases} \sum_{t=1}^{12} \hat{B}_i + \hat{A} \cdot \sum_{t=1}^{12} \sin\left(\frac{\pi}{6}t + \hat{\theta}\right) = \sum_{t=1}^{12} y_{it}, & dla \ \ i = 1, 2, ..., l; \\ \sum_{i=1}^{l} \sum_{t=1}^{12} \hat{B}_i \cos\left(\frac{\pi}{6}t + \hat{\theta}\right) + \frac{\hat{A}}{2} \cdot \sum_{i=1}^{l} \sum_{t=1}^{12} \sin\left(\frac{\pi}{3}t + 2\hat{\theta}\right) = \sum_{i=1}^{l} \sum_{t=1}^{12} y_{it} \cos\left(\frac{\pi}{6}t + \hat{\theta}\right); \\ \sum_{i=1}^{l} \sum_{t=1}^{12} \hat{B}_i \sin\left(\frac{\pi}{6}t + \hat{\theta}\right) + \hat{A} \cdot \sum_{i=1}^{l} \sum_{t=1}^{12} \sin^2\left(\frac{\pi}{6}t + \hat{\theta}\right) = \sum_{i=1}^{l} \sum_{t=1}^{12} y_{it} \sin\left(\frac{\pi}{6}t + \hat{\theta}\right). \end{cases} \qquad (5)$$

Korzystając z definicji sumy i zmieniając kolejność sumowania, w rezultacie otrzymujemy:

$$\begin{cases} \hat{B}_i \cdot 12 + \hat{A} \cdot \sum_{t=1}^{12} \sin\left(\frac{\pi}{6}t + \hat{\theta}\right) = \sum_{t=1}^{12} y_{it}, \quad dla \quad i = 1, 2, ..., l; \\ \sum_{i=1}^{l} \hat{B}_i \cdot \sum_{t=1}^{12} \cos\left(\frac{\pi}{6}t + \hat{\theta}\right) + \hat{A} \cdot \frac{l}{2} \cdot \sum_{t=1}^{12} \sin\left(\frac{\pi}{3}t + 2\hat{\theta}\right) = \sum_{i=1}^{l} \sum_{t=1}^{12} y_{it} \cos\left(\frac{\pi}{6}t + \hat{\theta}\right); \\ \sum_{i=1}^{l} \hat{B}_i \cdot \sum_{t=1}^{12} \sin\left(\frac{\pi}{6}t + \hat{\theta}\right) + \hat{A} \cdot l \cdot \sum_{t=1}^{12} \sin^2\left(\frac{\pi}{6}t + \hat{\theta}\right) = \sum_{i=1}^{l} \sum_{t=1}^{12} y_{it} \sin\left(\frac{\pi}{6}t + \hat{\theta}\right). \end{cases} \tag{6}$$

W pracy [5] udowodniono, że

$$\sum_{t=1}^{12} \sin\left(\frac{\pi}{6}t + \hat{\theta}\right) = \sum_{t=1}^{12} \cos\left(\frac{\pi}{6}t + \hat{\theta}\right) = \sum_{t=1}^{12} \sin\left(\frac{\pi}{3}t + 2\hat{\theta}\right) = 0, \text{ natomiast}$$

$$\sum_{t=1}^{12} \sin^2\left(\frac{\pi}{6}t + \hat{\theta}\right) = 6. \tag{7}$$

Podstawiając związki (7) do (6), po prostych przekształceniach ostatecznie otrzymujemy:

$$\begin{cases} \hat{B}_i + \frac{\hat{A}}{12} \cdot 0 = \frac{1}{12} \cdot \sum_{t=1}^{12} y_{it}, \quad dla \quad i = 1, 2, ..., l; \\ \frac{1}{l} \cdot \sum_{i=1}^{l} \hat{B}_i \cdot 0 + \frac{\hat{A}}{2} \cdot 0 = \frac{1}{l} \cdot \sum_{i=1}^{l} \sum_{t=1}^{12} y_{it} \cos\left(\frac{\pi}{6}t + \hat{\theta}\right); \\ \frac{1}{l} \cdot \sum_{i=1}^{l} \hat{B}_i \cdot 0 + \hat{A} \cdot 6 = \frac{1}{l} \cdot \sum_{i=1}^{l} \sum_{t=1}^{12} y_{it} \sin\left(\frac{\pi}{6}t + \hat{\theta}\right). \end{cases} \tag{8}$$

Z układu równań (8) obliczamy wartości ocen parametrów modelu (2):

$$\begin{cases} \hat{B}_i = \bar{y}_i, \quad dla \quad i = 1, 2, ..., l; \\ \hat{\theta} = \operatorname{arctg}\left[\sum_{t=1}^{12} \bar{y}_{\bullet t} \cos\frac{\pi}{6}t \Big/ \sum_{t=1}^{12} \bar{y}_{\bullet t} \sin\frac{\pi}{6}t\right]; \\ \hat{A} = \frac{1}{6}\left[\cos\hat{\theta} \cdot \sum_{t=1}^{12} \bar{y}_{\bullet t} \sin\frac{\pi}{6}t + \sin\hat{\theta} \cdot \sum_{t=1}^{12} \bar{y}_{\bullet t} \cos\frac{\pi}{6}t\right]; \end{cases} \tag{9}$$

w którym $\bar{y}_{i\bullet}$ jest średnią miesięczną i-tego roku, natomiast $\bar{y}_{\bullet t}$ jest średnią t--tego miesiąca z l lat (t = 1, 2,..., 12).

Łatwo zauważyć, że w modelu (2) trend wyraża się funkcją schodkową podaną w równaniu (9).

W przypadku, gdy lata mają jednakowy wpływ na opisywaną zmienną zależną, tzn. w modelu (2) $B_i = s$, wtedy jego oszacowanie z (9) wynosi:

$$\hat{B}_i = \hat{s} = \bar{y}_{..} = \frac{1}{12l} \sum_{i=1}^{l} \sum_{t=1}^{12} y_{it} \tag{10}$$

i nazywa się średnią globalną z danych empirycznych. Inaczej mówiąc, jest to średnia miesięczna, obliczona ze wszystkich lat, z których korzystaliśmy przy opracowywaniu modelu.

Model produkcji miesięcznej energii elektrycznej w Polsce

Racjonalna i skuteczna polityka energetyczna stanowi jedną z podstawowych przesłanek rozwoju społecznego i wzrostu ekonomicznego. Dążenie do oszczędnego gospodarowania energią jest obecnie oczywistą koniecznością, uzasadnioną pogłębiającym się wyczerpywaniem naturalnych zasobów energetycznych Ziemi i potrzebą zmniejszenia zanieczyszczenia środowiska naturalnego. Z wydobyciem i użytkowaniem każdego z surowców energetycznych wiąże się emisja niepożądanych substancji, które przyczyniają się do pogłębienia efektu cieplarnianego, zagrażającego zmianami klimatu. Najbardziej pożądana jest energia elektryczna. Wynika to z korzyści ekonomicznych, jakie daje jej stosowanie, oraz z komfortu, który zapewnia jej odbiorcom. Z tej przyczyny produkcja energii elektrycznej jest jednym z najszybciej rosnących elementów gospodarki rozwijających się państw. Stanowi ona zarazem jeden z najbardziej istotnych wskaźników obrazujących stopień ich rozwoju gospodarczego. Od niej też w głównej mierze zależy rozwój przemysłu, transportu, a nawet rolnictwa. Powszechnie uznanym wskaźnikiem zagospodarowania kraju jest zużycie energii elektrycznej na jednego mieszkańca. Natomiast o poziomie życia jego mieszkańców między innymi świadczy zużycie energii elektrycznej przez gospodarstwa domowe. Opiszemy i zbadamy kształtowanie się produkcji energii elektrycznej w Polsce, na podstawie danych miesięcznych z lat 1974–1993, podanych w tabeli 1 i zinterpretowanych graficznie na rysunku 1.

Aby oszacować parametry harmoniki, przytoczymy dane wyjściowe z tabeli 1 uwzględniające wyżej wymienione dwudziestolecie.

t	1	2	3	4	5	6	7	8	9	10	11	12
$\bar{y}_{\bullet t}$	12,07	11,12	11,48	10,18	9,45	8,83	8,83	9,04	9,46	10,93	11,53	12,23

Wykonujemy obliczenia pomocnicze:

$$\sum_{t=1}^{12} \bar{y}_{\bullet t} \cos\frac{\pi}{6}t = 9{,}422\,; \qquad \sum_{t=1}^{12} \bar{y}_{\bullet t} \sin\frac{\pi}{6}t = 3{,}752\,.$$

Wykorzystując wyprowadzone wzory (9), po zaokrągleniach ostatecznie otrzymujemy poszukiwany model:

$$\hat{y}_{it} = \bar{y}_{i\bullet} + 1{,}69 \sin\left(\frac{\pi}{6}t + 1{,}192\right) \quad \text{w TW} \cdot \text{h}, \quad (t = 1,2,...,12), \tag{11}$$

w którym $\bar{y}_{i\bullet}$ dla każdego $i = 1,2,...,20$ należy odczytać z tabeli 1.
Jego dopasowanie do danych empirycznych kształtuje się następująco: współczynnik zbieżności

$$\varphi^2 = \sum_{i=1}^{20}\sum_{t=1}^{12} (y_{it} - \hat{y}_{it})^2 \Big/ \sum_{i=1}^{20}\sum_{t=1}^{12} (y_{it} - \bar{y}_{\bullet\bullet})^2 = 66{,}9575/832{,}8316 = 0{,}0804\,;$$

współczynnik determinacji $R^2 = 1 - \varphi^2 = 0{,}9196$ – wnioskujemy z niego, że oszacowany model (11) tłumaczy około 92% zmienności badanej zmiennej zależnej; odchylenie standardowe reszt

$$s_e = \sqrt{\sum_{i=1}^{20}\sum_{t=1}^{12} \varepsilon_{it}^2 / (n-k)} = \sqrt{66{,}9575/(240-2)} = \\ = 0{,}56\ \text{TW} \cdot \text{h}\,;$$ współczynnik zmienności resztowej $v = 100\, s_e / \bar{y}_{\bullet\bullet} = 56/10{,}43 = 5{,}4\%$. Na rysunku 1 cienką linią zaznaczono wykres wyprowadzonej zależności (11). Używając tak prostych środków do opisu dominującej w tym zjawisku sezonowości, uzyskaliśmy dobrą zgodność modelu z danymi empirycznymi. Tym samym cel nasz został osiągnięty.

Podsumowanie

W zagadnieniach ekonomicznych przyszłość nie zawsze jest w pełni zdeterminowana przez stan aktualny i charakter podejmowanych działań. Z tego powodu wnioskowanie o przyszłym kształtowaniu się takich zjawisk nie jest niezawodne. Sprawą nauki jest opracowanie takich metod wnioskowania, aby obszar tej zawodności ograniczyć do minimum. Wieloletnie obserwacje różnych zjawisk upewniają nas, że przyszły stan większości zjawisk ekonomicznych i przyrodniczych zależy stochastycznie od ich stanu teraźniejszego oraz od przeszłego ich przebiegu w czasie. Stwarza to możliwość wykorzystania metod statystycznych do opisu i mierzenia zaobserwowanych prawidłowości rozwoju. Metody te nazywamy ekonometrycznym modelowaniem pewnego wycinka sfery zjawisk ekonomicznych. Ten sam problem można jednak rozwiązywać w różny sposób. Przyjęło się już, że budowanie modelu ekonometrycznego jest połączeniem nauki ze sztuką, podobnie jak w pracy architekta – ograniczonego wytycznymi ogólnymi zleceniodawcy. Propozycje rozwiązań mogą być różne, ale muszą bazować na metodach naukowych i mieć walory praktycznej użyteczności. W jednym artykule nie można rozwiązać wszystkich problemów związanych z tak trudnym zagadnieniem jak jednoczesna estymacja trendu i sezonowości. Dlatego problem ten rozwiązujemy etapami i przytaczamy praktyczne zastosowanie prezentowanych wzorów.

Literatura

BOX G.E.P., JENKINS G.M., 1983: *Analiza szeregów czasowych. Prognoza i sterowanie*. PWN, Warszawa. [1]

CHOW G.C., 1995: *Ekonometria*. PWN, Warszawa. [2]

Metody prognozowania. Zbiór zadań. 2000, red. B. RADZIKOWSKA, Wydawnictwo AE we Wrocławiu, Wrocław. [3]

Prognozowanie gospodarcze. Metody i zastosowania. 1997. Red. naukowa M. CIEŚLAK, PWN, Warszawa. [4]

SMOLIK S., 1995: *Uproszczona procedura estymacji modelu wahań okresowych*. „Przegląd Statystyczny", R. XLII, z. 3–4, s. 449–457. [5]

SMOLIK S., 1998: *Oszczędne modele dla okresowych szeregów czasowych*. Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodar-

czych. Red. A. ZELIAŚ, Akademia Ekonomiczna w Krakowie, s. 177–188. [6]
SMOLIK S., 1997: *Sezonowość w opisie procesów rolniczych.* „Wiadomości Statystyczne", nr 4, s. 10–14. [7]
STECZKOWSKI J., ZELIAŚ A., 1982: *Analiza wariancyjna i kowariancyjna w badaniach ekonomicznych.* PWN, Warszawa. [8]
ZELIAŚ A., 1997: *Teoria prognozy.* Wyd. 3 zmienione, PWE, Warszawa. [9]
ZIELIŃSKI Z., 1969: *Ekonometryczne metody analizy wahań sezonowych.* Zeszyty Naukowe Politechniki Szczecińskiej, nr 112 (rozprawa habilitacyjna). [10]

Statistic modelling of trend-based and season-related phenomena

Abstract

In quite a considerable number of economic and natural issues, during research, observation series are obtained with a very distinct seasonal dependence and trend, whether constant or vaguely outlined. Chance variations (randomness) of this type is encountered in climatology, when it comes to describing average monthly temperatures of air and the ground, and in temporal series describing production of, not infrequently elementary, commodities. For description of these phenomena, the following model has been proposed:

$$y_{it} = B_i + A\sin(\pi t/6 + \theta) + \varepsilon_{it}, \text{ for } i = 1, 2, ..., l \text{ years; } t = 1, 2, ..., 12 \text{ months.}$$

It has been proved that evaluation of its parameters with the use of the least square (minimum chi-square) method, based on the empirical data table with l lines and 12 columns, assumes the following form:

$\hat{B}_i = \bar{y}_{i\bullet}$, for each $i = 1, 2, ..., l$; whereas $\bar{y}_{i\bullet}$ is the monthly average of an i-th year,

$$\hat{\theta} = \operatorname{arctg}\left[\sum_{t=1}^{12} \bar{y}_{\bullet t} \cos\frac{\pi}{6}t \Big/ \sum_{t=1}^{12} \bar{y}_{\bullet t} \sin\frac{\pi}{6}t\right],$$

$$\hat{A} = \frac{1}{6}\left[\cos\hat{\theta} \cdot \sum_{t=1}^{12} \bar{y}_{\bullet t} \sin\frac{\pi}{6}t + \sin\hat{\theta} \cdot \sum_{t=1}^{12} \bar{y}_{\bullet t} \cos\frac{\pi}{6}t\right],$$

whereas $\bar{y}_{\bullet t}$ is the average of a t-th month $(t = 1, 2, ..., 12)$, calculated basing on l years.

The thus proposed model has been used in a description of power (electric energy) production in Poland, employing monthly empirical data from the period of 1974–1993.