

Organizatorzy seminarium, programując obrady, bez wątplenia chcieli powiązać historię polskiej statystyki ze współczesnością, wykazać niezbędność i użyteczność badań statystycznych, a zarazem podkreślić wielką odpowiedzialność, która spoczywa na barkach ich realizatorów. Świadczy o tym chociażby dobór tematyki referatów i ranga ich autorów, a także fakt, że obrady odbywają się niemal w przeddzień wielkich operacji badawczych, jakimi są *Powszechny Spis Rolny 2010* oraz *Narodowy Spis Powszechny Ludności i Mieszkań 2011*. Jestem przekonany, że przesłania płynące z treści zgłoszonych tutaj opracowań i dyskusji nad nimi dotrą do właściwych adresatów, a wzmocni je powaga miejsca obrad i zbiorowy głos uczestników seminarium.

Dzisiejsze seminarium zgromadziło liczne grono przedstawicieli społeczności statystyków polskich. Przybyli na nie naukowcy z różnych ośrodków akademickich, pracownicy służb statystyki publicznej, członkowie PTS i osoby nienależące do tej organizacji, profesjonalisci z wielu dziedzin, czynni zawodowo lub realizujący się w innej formie aktywności. Liczebność i skład tego audytorium dają podstawy do stwierdzenia, że ustanowienie Dnia Statystyki Polskiej było dobrą inicjatywą, która rozwinie się w przyszłości i przyniesie wiele pożytku.

W imieniu własnym i Rady Głównej PTS pragnę gorąco podziękować Panu Prezesowi GUS — prof. drowi Józefowi Oleńskiemu — za zaproszenie na dzisiejsze seminarium naukowe. Jestem przekonany, że to świetnie przygotowane spotkanie będzie dobrym początkiem nowej tradycji związanej z obchodami Dnia Statystyki Polskiej.

Całej społeczności statystyków przekazuję życzenia dalszych sukcesów i satysfakcji z pełnionej misji, a zebrany dzisiaj w Sali Kolumnowej Sejmu RP życzę owocnych obrad, nawiązania nowych kontaktów i wzmocnienia starych więzi.

## STUDIA METODOLOGICZNE

**Natalia NEHREBECKA, Sylwia GRUDKOWSKA**

### Wykorzystanie metody *epsilon* do badania wpływu czynników determinujących opinie konsumentów

---

Wskaźniki koniunktury konsumenckiej dostarczają informacji o subiektywnej ocenie zjawisk ekonomicznych przez respondentów. Najczęściej są one zbudowane na podstawie badań opinii na temat bieżącej i przyszłej sytuacji ekonomicznej kraju i gospodarstw domowych, pozycji na rynku pracy oraz planów inwestycyjnych. Pomimo że dane, których źródłem są tego typu badania, mają

charakter subiektywny, stanowią cenne źródło informacji o kierunku rozwoju procesów ekonomicznych w gospodarce. Przykładowo, oczekiwania płacowe znalazły zastosowanie w teorii cyklu życia w charakterze predyktorów wydatków gospodarstw domowych. Opinie i prognozy konsumenckie mają również istotne znaczenie dla kształtowania się cykli koniunkturalnych. Zostały one włączone przez Samuelsona do modelu wyjaśniającego bieżący poziom konsumpcji za pomocą oczekiwań co do wysokości przyszłego dochodu (Westerhoff, 2006).

Badania empiryczne wskazują, że kształtowanie się prognoz dokonywanych przez konsumentów ma w dużej mierze charakter ekstrapolacji bieżących i przeszłych wartości. W szczególności dotyczy to oczekiwań inflacyjnych. Przeprowadzone badania dotyczące strefy euro (Forsells, Kenny, 2002) wskazują, że oczekiwania inflacyjne są nieobciążonym estymatorem przyszłego kierunku rozwoju inflacji. Autorzy stwierdzili również, że predykcje nie są w pełni efektywne, gdyż błędy prognoz nie mają charakteru losowego i mogą być objaśnione przy pomocy informacji o podaży pieniądza oraz wysokości stóp procentowych. Oznacza to, że wpływ innych zmiennych na przewidywaną inflację nie jest zanedbywany.

Celem artykułu jest zbadanie wpływu niektórych zmiennych makroekonomicznych na formułowanie prognoz konsumenckich. Oszacowana zostanie również wielkość udziału poszczególnych zmiennych w wyjaśnieniu danego wskaźnika koniunktury. Uzyskane wyniki posłużą do weryfikacji hipotezy, że każdy ze wskaźników koniunktury kształtuje się pod dominującym wpływem innego zestawu zmiennych.

### *KORELACJA ZMIENNYCH A EFEKTY BEZPOŚREDNIE, CAŁKOWITE I CZĘŚCIOWE*

Charakterystycznym problemem modeli ekonomicznych jest wysoka korelacja zmiennych objaśniających. Jest ona spowodowana współzależnością zjawisk ekonomicznych wynikającą z faktu, że na kształtowanie się regresorów wybranych do modelu często wpływają te same zjawiska. W wyniku tego zmienne objaśniane wykazują zbliżony trend. W przypadku silnego skorelowania zmiennych określenie wkładu każdej z nich w wyjaśnienie zmiennej zależnej jest problematyczne. Utrudnione jest wówczas wybranie zbioru regresorów, których wkład w predykcję badanej wielkości jest znaczący. Problem ten stanowi poważną przeszkodę w analizie mechanizmów ekonomicznych.

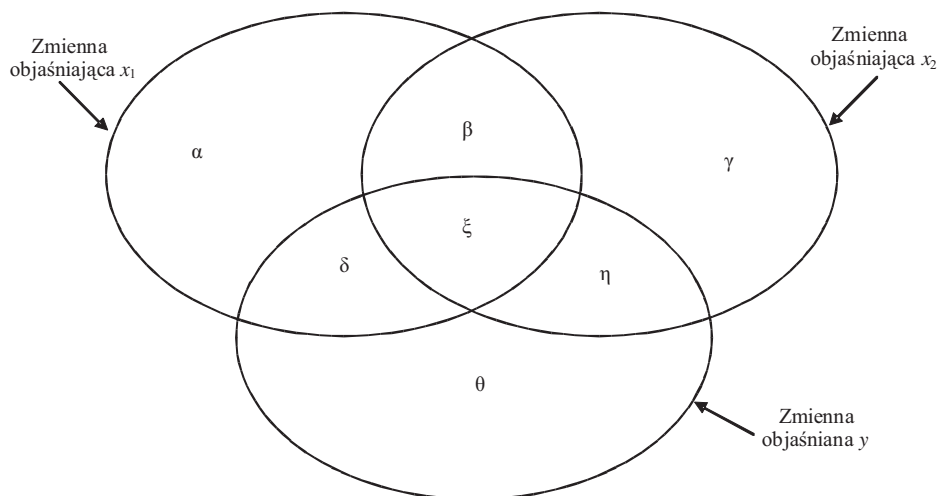
Obecnie istnieje kilka sposobów szacowania relatywnego wpływu regresorów na zmienną zależną. Większość z nich nie uwzględnia w sposób kompletny interakcji pomiędzy zmiennymi, w wyniku czego część informacji dotycząca korelacji jest przez nie pomijana.

Optymalna metoda powinna prowadzić do obliczenia wskaźnika relatywnej siły regresora (Johnson, 2000). Jest on zdefiniowany jako wkład zmiennej niezależnej w estymację zmiennej objaśnianej, uwzględniający zarówno bezpośredni wpływ tej zmiennej na kształtowanie się badanego zjawiska, jak i jej pośrednie

działanie w połączeniu z innymi zmiennymi objaśniającymi. Algorytmem spełniającym ten warunek jest m.in. metoda *epsilon*.

W przypadku zmiennych ortogonalnych względny wkład każdego regresora może być zdefiniowany jako stosunek współczynnika korelacji pomiędzy jednym z tych regresorów a zmienną objaśnianą do kwadratu współczynnika korelacji wielorakiej wszystkich zmiennych modelu. W typowych sytuacjach nie można jednak uznać, że regresory są od siebie niezależne. Na wykresie 1 przedstawiono schemat powiązań pomiędzy zmiennymi niezależnymi a zmienną zależną w modelu o postaci  $y = \mathbf{X}\beta + \varepsilon$ , gdzie  $y$  jest wektorem obserwacji zmiennej objaśnianej,  $\mathbf{X}$  jest macierzą obserwacji dwóch zmiennych objaśniających:  $x_1$  i  $x_2$ ,  $\beta$  jest wektorem nieznanych parametrów,  $\varepsilon$  — wektorem zaburzeń losowych.

Wykr. 1. ZALEŻNOŚCI POMIĘDZY WARIANCJAMI ZMIENNYCH W MODELU Z DWOMA REGRESORAMI



Źródło: opracowanie własne.

Na wykresie zobrazowano wariancję każdej ze zmiennych modelu. Suma składowych  $\alpha$ ,  $\beta$ ,  $\delta$  i  $\zeta$  oznacza wariancję zmiennej  $x_1$ . Wariancja zmiennej  $x_2$  jest równa sumie wartości  $\gamma$ ,  $\beta$ ,  $\eta$  i  $\zeta$ , natomiast wariancja  $y$  jest równa sumie

$\delta, \theta, \eta$  i  $\zeta$ . Wartości  $\delta$  oraz  $\eta$  to części wariancji  $y$  wyjaśnione indywidualnie przez odpowiednio  $x_1$  i  $x_2$ . Wariancja  $y$  wyjaśniona wspólnie przez  $x_1$  i  $x_2$  wynosi  $\zeta$ .

Przedstawiony schemat można wykorzystać do ilustracji wskaźników relatywnego wpływu wyróżnionych przez Lebreton i in. (2004). Pierwszy z nich jest oparty na bezpośrednim wpływie regresorów  $x_1$  i  $x_2$  na  $y$  mierzonym odpowiednio sumą wielkości  $\delta$  i  $\zeta$  oraz sumą  $\eta$  i  $\zeta$ . Wynika stąd, że w tym mierniku ta część wariancji, która jest tłumaczona przez oba zjawiska, została przyporządkowana do każdego z tych regresorów. Wskaźnik oparty na efektach bezpośrednich jest zatem w pełni miarodajny jedynie w przypadku zmiennych nieskorelowanych.

Efekt całkowity oznacza wkład danego predykatora do wyjaśnienia zjawiska przy obecności pozostałych zmiennych. Jest on najczęściej przedstawiany za pomocą kwadratu wystandaryzowanych<sup>1</sup> współczynników regresji z modelu ( $\beta^2$ ). Relatywna siła regresora obliczona na podstawie efektów całkowitych nie bierze jednakże pod uwagę oddziaływania regresora w izolacji od innych zmiennych, gdyż pomija korelację pomiędzy zmiennymi.

Dokonanie oceny relatywnego wkładu danej zmiennej do  $R^2$  modelu z wyłączeniem wpływu pozostałych regresorów jest możliwe za pomocą wskaźnika efektów częściowych. Mierzy on siłę związku pomiędzy zmienną objaśnianą a wybranym regresorem, z pominięciem oddziaływania pozostałych zmiennych objaśniających. W przypadku omawianego modelu wskaźnik ten obliczony dla  $y$  i  $x_1$  przyjmuje postać:

$$r^2_{x_1y, x_2} = \left( \frac{r_{x_1y} - r_{x_1x_2}r_{yx_2}}{\sqrt{(1 - r_{x_1x_2})(1 - r_{yx_2})}} \right)^2$$

Wadą współczynnika kwadratu korelacji częściowej jest brak podziału części wariancji wspólnie wyjaśnianej przez regresory pomiędzy  $x_1$  i  $x_2$ . Problem ten można zobrazować posługując się wykr. 1. Efekty częściowe dla  $x_1$  i  $x_2$  są obliczane odpowiednio jako:  $\frac{\delta}{\delta + \theta + \eta + \zeta}$  i  $\frac{\eta}{\delta + \theta + \eta + \zeta}$ . Wynika stąd, że wartość  $\zeta$  nie jest dzielona pomiędzy predykatory.

Każdy ze wskaźników oparty tylko na jednym z tych efektów nie wykorzystuje pełnej informacji o zależnościach pomiędzy zmiennymi. Z tego względu w opracowaniu zdecydowano się na wykorzystanie metody *epsilon*, która pozwala na uwzględnienie pełnej wiedzy o korelacji pomiędzy zmiennymi w modelu.

---

<sup>1</sup> Współczynniki wystandaryzowane to współczynniki obliczone dla zmiennych, których wariancja jest równa 1.

## METODA EPSILON

Problem korelacji pomiędzy zmiennymi, który był główną przyczyną niedoskonałości wcześniejszych metod ewaluacji siły predyktora, został w metodzie *epsilon* usunięty poprzez przekształcenie zbioru regresorów do zmiennych ortogonalnych. W wyniku tego możliwe jest wykorzystanie współczynników regresji do oszacowania wkładu danego regresora do  $R^2$  modelu.

Pierwszym krokiem algorytmu ortogonalizacji (Johnson, 2000) jest dekompozycja macierzy regresorów  $\mathbf{X}$  według wartości szczególnych<sup>2</sup>:

$$\mathbf{X} = \mathbf{P} \Delta \mathbf{Q}^T$$

gdzie:

$\mathbf{X}$  — macierz  $n \times p$ ,

$\mathbf{P}$  — macierz  $n \times n$  złożona z wektorów własnych macierzy  $\mathbf{X} \mathbf{X}^T$ ,

$\mathbf{Q}$  — macierz  $p \times p$  złożona z wektorów własnych macierzy  $\mathbf{X}^T \mathbf{X}$ ,

$T$  — transpozycja macierzy,

$\Delta$  — macierz  $p \times p$  diagonalna, w której na głównej przekątnej znajdują się pierwiastki kwadratowe wartości własnych macierzy  $\mathbf{X} \mathbf{X}^T$  i  $\mathbf{X}^T \mathbf{X}$ .

Jeżeli w macierzy  $\mathbf{X}$  nie istnieją w pełni współliniowe regresory<sup>3</sup> (Johnson, 1966), to najlepsza ortogonalna aproksymacja  $\mathbf{X}$  ma postać:

$$\mathbf{Z} = \mathbf{P} \mathbf{Q}^T$$

gdzie  $\mathbf{Z}$  jest macierzą  $n \times p$ .

Taki wybór macierzy  $\mathbf{Z}$  minimalizuje sumę kwadratów reszt pomiędzy zmiennymi znajdującymi się w macierzach  $\mathbf{X}$  i  $\mathbf{Z}$ . Względny wkład  $k$ -tej zmiennej znajdującej się w macierzy  $\mathbf{Z}$ , oznaczonej jako  $z_k$ , na  $j$ -tą zmienną z macierzy  $\mathbf{X}$ , oznaczonej przez  $x_j$ , jest równy kwadratowi wystandaryzowanego współczynnika regresji  $\lambda_{jk}^*$ . Ze względu na brak korelacji pomiędzy

---

<sup>2</sup> Każda macierz  $m \times n$  może być przedstawiona jako iloczyn macierzy  $\mathbf{P}$  o wymiarze  $m \times n$ , której kolumny składają się ze wzajemnie prostopadłych wektorów, diagonalnej macierzy kwadratowej  $\Delta$  o rozmiarze  $n \times n$  o nieujemnych wartościach na diagonalu oraz transpozycji macierzy  $\mathbf{Q}$  o wymiarze  $n \times n$ . To zagadnienie omawia szczegółowo Wall i in. (2003).

<sup>3</sup> Jeśli warunek ten jest spełniony, to wszystkie wartości własne  $\mathbf{X} \mathbf{X}^T$  i  $\mathbf{X}^T \mathbf{X}$  są różne od zera.

zmieniami zastosowanie wskaźnika opartego na efektach całkowitych nie powoduje tu wystąpienia problemu nieuwzględniania części korelacji. Współczynnik  $\lambda_{jk}^*$  jest elementem macierzy  $\Lambda^*$  obliczanej z regresji  $\mathbf{Z}$  na  $\mathbf{X}$ :

$$\Lambda^* = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{X} = (\mathbf{Q} \mathbf{P}^T \mathbf{P} \mathbf{Q}^T)^{-1} \mathbf{Q} \mathbf{P}^T \mathbf{P} \Delta \mathbf{Q}^T = \mathbf{I}^{-1} \mathbf{Q} \Delta \mathbf{Q}^T = \mathbf{Q} \Delta \mathbf{Q}^T$$

Ponieważ  $\lambda_{jk}^*$  są współczynnikami korelacji, to oznaczają one zarówno część wariancji  $z_k$  wyjaśnioną przez  $x_j$ , jak i część wariancji  $x_j$  wyjaśnioną przez  $z_k$ .

$$\begin{aligned} \text{Ponieważ } \mathbf{Z}^T \mathbf{X} &= \mathbf{Q} \mathbf{P}^T \mathbf{P} \Delta \mathbf{Q}^T = \mathbf{I}^{-1} \mathbf{Q} \Delta \mathbf{Q}^T = \mathbf{Q} \Delta \mathbf{Q}^T \text{ i} \\ \mathbf{X}^T \mathbf{Z} &= \mathbf{Q} \Delta \mathbf{P}^T \mathbf{P} \mathbf{Q}^T = \mathbf{I}^{-1} \mathbf{Q} \Delta \mathbf{Q}^T = \mathbf{Q} \Delta \mathbf{Q}^T, \end{aligned}$$

to zachodzi następująca równość:

$$\mathbf{Z}^T \mathbf{X} = \mathbf{X}^T \mathbf{Z}$$

Wkład zmiennej  $x_j$  w wyjaśnienie  $y$  może być zatem przybliżony wartością:  $\sum_k \lambda_{jk}^* \beta_k^*$ , gdzie  $\beta_k^*$  jest współczynnikiem regresji liniowej  $y$  na  $z_k$ .

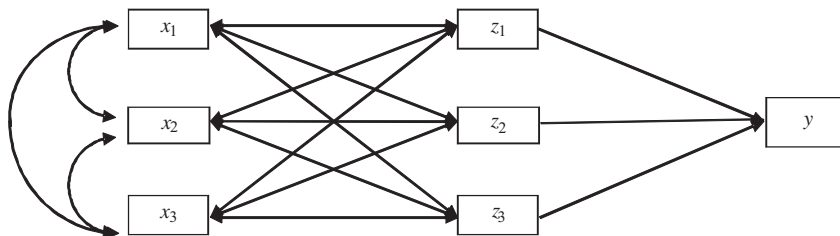
Wektor współczynników  $\beta^*$  jest obliczany jako:

$$\beta^* = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T y = (\mathbf{Q} \mathbf{P}^T \mathbf{P} \mathbf{Q}^T)^{-1} \mathbf{Q} \mathbf{P}^T y = \mathbf{I}^{-1} \mathbf{Q} \mathbf{P}^T y = \mathbf{Q} \mathbf{P}^T y$$

Zastosowanie tej ortogonalizacji pierwotnych zmiennych pozwala na przedstawienie każdej ze zmiennych  $z_1, \dots, z_p$  jako liniowej kombinacji regresorów  $x_1, \dots, x_p$ . Zabieg ten umożliwia rozdystrybuowanie wariancji zmiennej  $y$  wyjaśnionej zmienną  $z_k$  na poszczególne regresory  $x_1, \dots, x_p$  proporcjonalnie do wariancji  $z_k$  wyjaśnionej zmienną  $x_j$ . Schemat zależności pomiędzy zmiennymi  $z_1, \dots, z_p$ ,  $x_1, \dots, x_p$  oraz  $y$  dla przypadku trzech zmiennych  $x_1, x_2, x_3$  i trzech zmiennych  $z_1, z_2, z_3$  przedstawia wyk. 2.

Przedstawiony algorytm metody *epsilon* prowadzi do otrzymania relatywnych wag przypisanych do każdej z oryginalnych zmiennych modelu. Wagi te mogą być interpretowane jako relatywny wkład zmiennej, do której się odnoszą, w wyjaśnienie zachowania zmiennej  $y$ , gdyż ich suma jest równa  $R^2$  modelu. W artykule wagi wynikowe zostały przekształcone tak, aby ich suma stanowiła 100%.

Wykr. 2. SCHEMAT POWIĄZAŃ POMIĘDZY ZMIENNYMI W METODZIE *EPSILON*



Źródło: opracowanie własne.

### *DETERMINANTY SUBIEKTYWNEJ OCENY ZJAWISK EKONOMICZNYCH PRZEZ KONSUMENTÓW*

Analizie poddano sześć wskaźników ufności konsumenckiej utworzonych na podstawie wyników badań postaw konsumentów przeprowadzanych przez GUS z częstotliwością miesięczną<sup>4</sup>. Próba obejmuje okres styczeń 2004 r.—sierpień 2008 r. Każdy ze wskaźników jest wyliczany jako suma iloczynów frakcji każdego wariantu odpowiedzi na odpowiednie pytanie i przyporządkowanej temu wariantowi wagi.

Kształtowanie się każdego ze wskaźników w okresie objętym próbą przedstawiono na wykr. 3—8.

Ze względu na brak dostępności danych o jednostkowej charakterystyce respondentów, takich jak: wiek, płeć, wykształcenie czy dochody, zbiór potencjalnych zmiennych objaśniających ograniczono do danych makroekonomicznych.

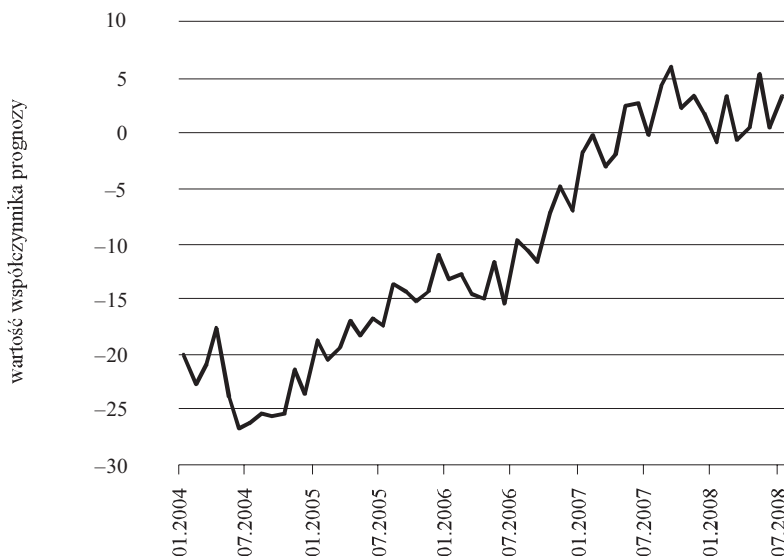
Wybrany przez autorki zbiór czynników, które mogą wpływać na percepcję konsumentów w obszarach, których dotyczyły pytania obejmuje:

- kursy walutowe dolara, euro i franka szwajcarskiego (USD/PLN, EUR/PLN, CHF/PLN),
- realne oprocentowanie złotych kredytów udzielonych gospodarstwom domowym na cele mieszkaniowe w %,
- realne oprocentowanie złotych kredytów udzielonych gospodarstwom domowym na cele konsumpcyjne w %,

<sup>4</sup> Szczegółowe informacje dotyczące metodologii badań nad koniunkturą są dostępne w materiale *Koniunktura konsumencka. Tendencje zmian październik 2003—styczeń 2004*. W analizie pominięto Wyprzedzający Wskaźnik Ufności Konsumenckiej i Bieżący Wskaźnik Ufności Konsumenckiej, gdyż ich wartości są wyliczane na podstawie pozostałych wskaźników.

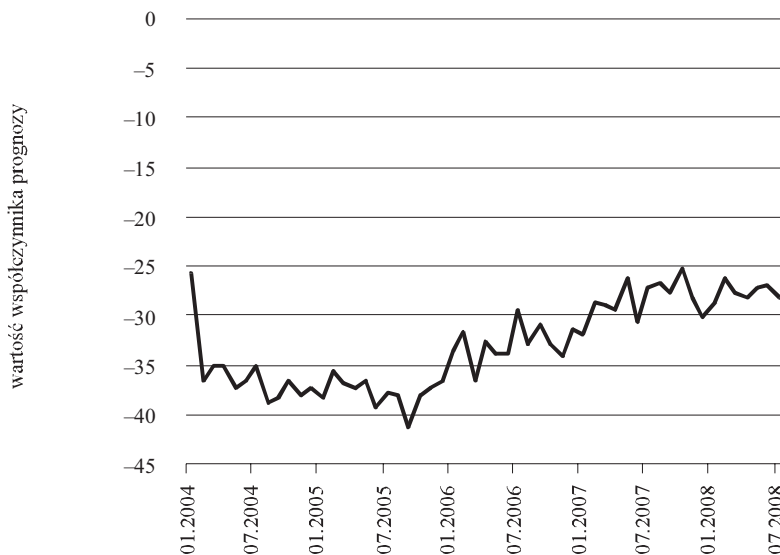
- stopę bezrobocia w %,
- niewykorzystane dłużej niż miesiąc oferty pracy według stanu na koniec miesiąca w tys.,
- wskaźnik cen paliw (analogiczny okres poprzedniego roku = 100),
- wskaźnik cen żywności i napojów bezalkoholowych (analogiczny okres poprzedniego roku = 100),
- liczbę mieszkań oddanych do użytku, w tym mieszkań indywidualnych oraz przeznaczonych na sprzedaż lub wynajem,
- obroty w handlu detalicznym (2000=100),
- przeciętne wynagrodzenia miesięczne brutto bez wypłat z zysku w sektorze przedsiębiorstw w zł,
- wartość kredytów i pożyczek mieszkaniowych udzielonych osobom prywatnym (analogiczny okres poprzedniego roku = 100),
- wartość kredytów i pożyczek w sprzedaży ratalnej udzielonych osobom prywatnym (analogiczny okres poprzedniego roku = 100),
- wartość kredytów i pożyczek w rachunku bieżącym i związanych z funkcjonowaniem kart płatniczych udzielonych osobom prywatnym (analogiczny okres poprzedniego roku = 100),

Wykr. 3. OCENA MOŻLIWOŚCI DOKONYWANIA WAŻNYCH ZAKUPÓW



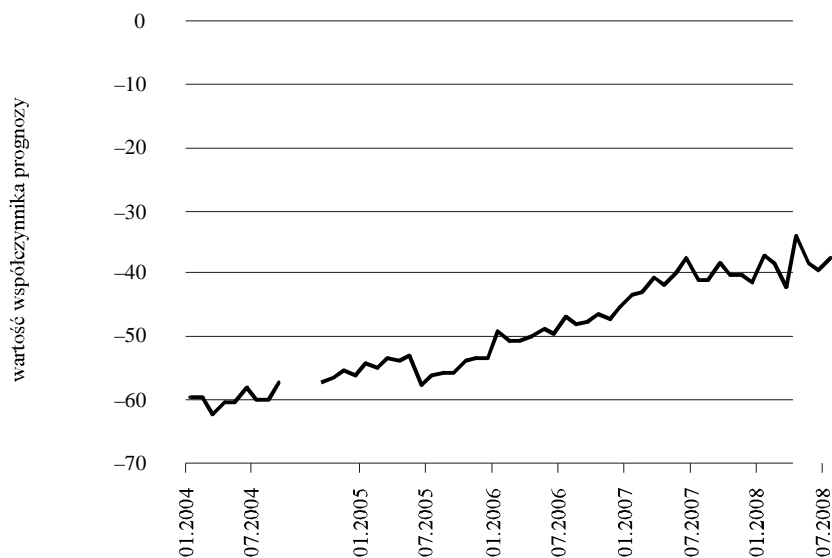
Źródło: opracowanie własne.

**Wykr. 4. PROGNOZA NA KOLEJNE 12 MIESIĘCY DOKONYWANIA WAŻNYCH ZAKUPÓW**



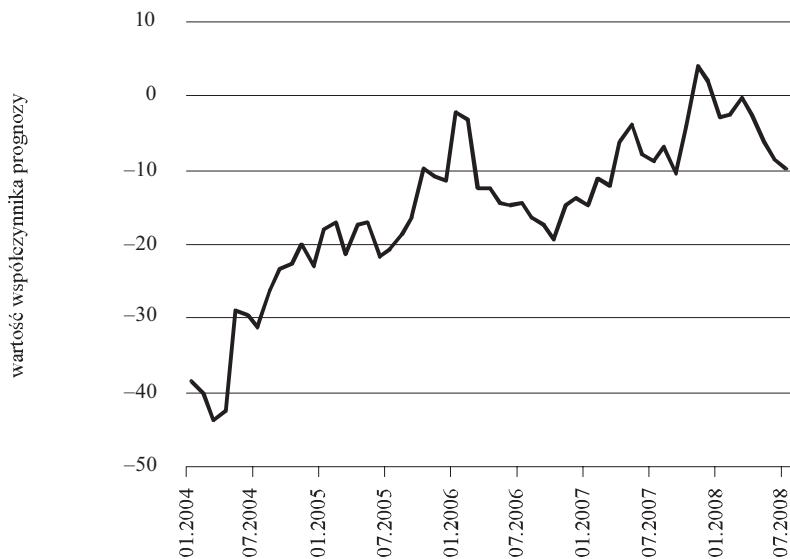
Źródło: opracowanie własne.

**Wykr. 5. PROGNOZA NA KOLEJNE 12 MIESIĘCY OSZCZĘDZANIA PIENIĘDZY**



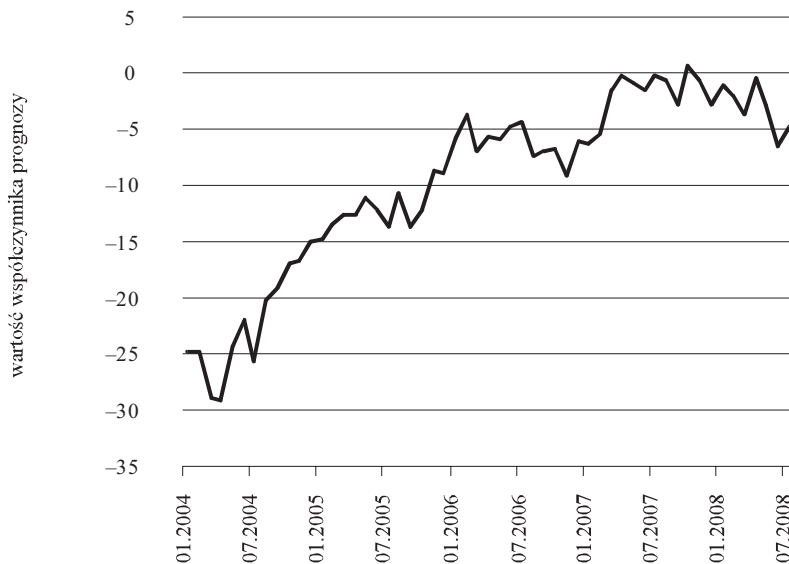
Źródło: opracowanie własne.

**Wykr. 6. PROGNOZA NA KOLEJNE 12 MIESIĘCY ZMIANY OGÓLNEJ SYTUACJI EKONOMICZNEJ W KRAJU**



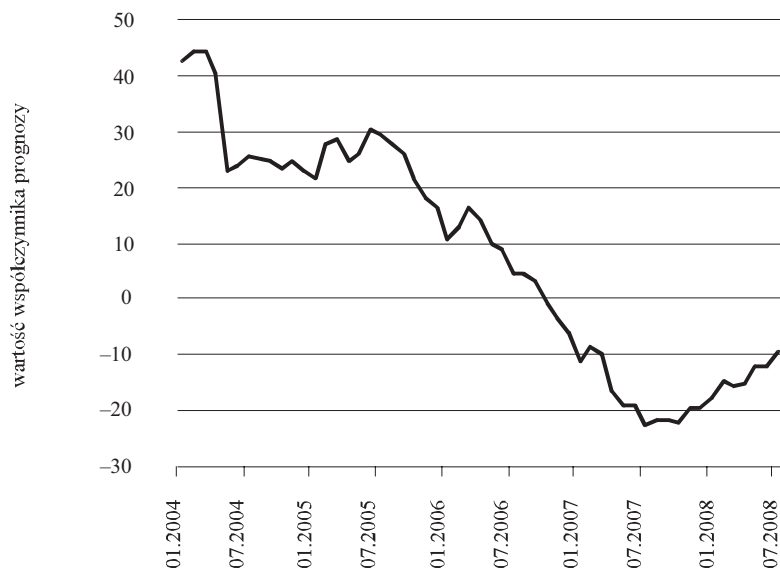
Źródło: opracowanie własne.

**Wykr. 7. PROGNOZA NA KOLEJNE 12 MIESIĘCY ZMIANY SYTUACJI FINANSOWEJ GOSPODARSTWA DOMOWEGO**



Źródło: opracowanie własne.

Wykr. 8. PROGNOZA NA KOLEJNE 12 MIESIĘCY ZMIANY POZIOMU BEZROBOCIA



Źródło: opracowanie własne.

Wskaźniki zostały dobrane tak, aby reprezentować najistotniejsze z punktu widzenia konsumenta zjawiska ekonomiczne, przy uwzględnieniu dostępności szeregów czasowych o odpowiedniej długości i częstotliwości.

Regresory związane z budownictwem mieszkaniowym i obrotami w handlu detalicznym poddano odsezonowaniu metodą TRAMO/SEATS<sup>5</sup> ze względu na bardzo wyraźną obecność w nich efektów sezonowych zaburzających właściwą ocenę relacji pomiędzy zmiennymi.

### WYNIKI

Otrzymane wyniki (tabl.) wskazują, że największe znaczenie dla możliwości dokonywania ważnych zakupów mają sygnały z rynku pracy. Na ocenę respondentów możliwości dokonywania zakupów dóbr trwałego użytku w niemal równym stopniu wpływa wysokość stopy bezrobocia i liczba niewykorzystanych ofert pracy.

Na obecną ocenę możliwości dokonywania ważnych zakupów przez respondentów silnie oddziałuje wskaźnik obrotów w handlu detalicznym, natomiast znaczenie wskaźników cen jest marginalne. Oznacza to, że konsumenci udzielając odpowiedzi kierują się realnymi możliwościami zakupu, a nie poziomem cen

<sup>5</sup> Algorytm TRAMO/SEATS jest opisany m.in. w A. Maravall (2006).

towarów. Hipotezę tę potwierdza również znaczny udział (8,2%) przeciętnego wynagrodzenia w wyjaśnieniu  $R^2$  modelu.

Relatywnie duże znaczenie ma wartość kredytów, zarówno w sprzedaży ratalnej jak i mieszkaniowych. Jest to sygnał wskazujący, że istotna część ważnych zakupów jest finansowana kredytami. Wpływ oprocentowania kredytów jest niewielki, gdyż w okresie próby obciążenie gospodarstw domowych kosztami spłaty zadłużenia było stosunkowo niewielkie. Obciążenie budżetów kredytami mieszkaniowymi uwidacznia się w wysokim udziale kursu franka szwajcarskiego (9,4%). Wysokość rat kredytów mieszkaniowych, których spłata jest dla gospodarstw domowych kwestią priorytetową, determinuje możliwości poniesienia kosztów tych zakupów.

Kursy pozostałych walut mają mniejsze znaczenie, co jest zgodne z obserwowaną w gospodarce polskiej tendencją, że zakupy tych dóbr są finansowane kredytami złotowymi<sup>6</sup> lub gotówką. Waga wskaźnika cen paliw jest ponad dwukrotnie wyższa niż wskaźnika cen towarów i usług konsumpcyjnych. Wynika stąd, że znaczenie przypisywane przez konsumentów cenom paliw jest nadmierne. Z badań budżetów gospodarstw domowych wynika bowiem, że wydatki związane z eksploatacją prywatnych środków transportu stanowią niecałe 6% ogółu wydatków na towary i usługi konsumpcyjne, podczas gdy analogiczny udział żywności wynosi prawie 30%.

Podstawowe znaczenie dla prognozy dokonywania ważnych zakupów ma wartość kredytów i pożyczek mieszkaniowych oraz liczba oddanych do użytku mieszkań indywidualnych. Wpływ tych regresorów może być dwojaki. Wzrost wartości kredytów na nieruchomości wskazuje na poprawę sytuacji finansowej konsumentów i w rezultacie zwiększenie ich możliwości nabywczych. Powoduje on jednak ograniczenie poważniejszych wydatków przez osoby, które posiadają kredyt hipoteczny. Stanowisko to potwierdza wysoki udział liczby mieszkań indywidualnych oddanych do użytku w wyjaśnieniu zachowania omawianej zmiennej. Sytuacja na rynku mieszkaniowym ma zatem znaczenie dla możliwości ponoszenia wydatków na dobra trwałego użytku.

Na przyszłe wydatki, w podobnym stopniu jak na bieżące, wpływa indeks obrotów w handlu detalicznym. Klienci przewidują możliwości dokonania zakupów na podstawie bieżącej koniunktury w handlu. Prognozy dokonywania ważnych zakupów mają zatem charakter adaptacyjny, aczkolwiek skala wpływu części pozostałych czynników jest równie wysoka.

Uzyskane wyniki wskazują, że ocena możliwości oszczędzania pieniędzy jest zdeterminowana głównie przez informacje o kształtowaniu się kursu walutowego franka szwajcarskiego. Ponownie wskazuje to na dominujący wpływ grupy osób obciążonych kredytami mieszkaniowymi na wartość tego wskaźnika koniunktury. Zgodnie z oczekiwaniami, poważny wpływ na potencjalne oszczęd-

---

<sup>6</sup> Według danych NBP ponad 90% wartości kredytów w 2008 r. w systemie sprzedaży ratalnej stanowią kredyty złotowe.

ności mają obroty w handlu detalicznym. Ograniczenie bieżących zakupów w sposób oczywisty przekłada się bowiem na przyszłe oszczędności.

#### ZESTAWIENIE RELATYWNYCH WKŁADÓW REGRESORÓW DO FORMOWANIA OPINII RESPONDENTÓW

Regresory	Ocena możliwości dokonywania ważnych zakupów	Prognoza na kolejne 12 miesięcy				
		dokonywania ważniejszych zakupów	oszczędzania pieniędzy	zmiany ogólnej sytuacji ekonomicznej w kraju	zmiany poziomu bezrobocia	zmiany sytuacji finansowej gospodarstwa domowego
relatywny udział w $R^2$ w %						
Stopa bezrobocia .....	9,5	8,0	9,9	7,8	9,5	8,6
Oferty pracy na koniec miesiąca .....	9,1	8,7	8,9	5,8	9,4	7,8
CHF — średnioważony kurs walutowy sprzedaży w PLN .....	9,4	7,5	11,0	15,2	9,5	14,2
EUR — średnioważony kurs walutowy w PLN .....	3,0	5,5	4,8	14,2	5,3	10,8
USD — średnioważony kurs walutowy w PLN .....	2,7	3,1	3,2	2,9	2,5	2,7
Kredyty i pożyczki w sprzedaży ratalnej udzielone osobom prywatnym .....	6,1	3,8	5,4	11,3	4,2	9,9
Kredyty i pożyczki mieszkaniowe udzielone osobom prywatnym .....	6,1	10,1	7,2	5,2	10,4	6,7
Kredyty w rachunku bieżącym i związane z funkcjonowaniem kart kredytowych .....	9,1	8,1	9,4	8,1	8,3	8,0
Realne oprocentowanie kredytów konsumpcyjnych ogółem .....	3,5	1,8	1,5	1,6	2,1	2,1
Realne oprocentowanie kredytów mieszkaniowych ogółem .....	4,4	3,9	4,3	3,1	5,9	2,9
Mieszkania oddane do użytku:						
indywidualne .....	5,6	10,1	3,7	1,4	3,1	1,5
na sprzedaż lub wynajem .....	7,3	4,9	6,6	6,3	5,8	5,4
pozostałe .....	0,4	0,5	0,8	0,7	1,3	1,0
Obroty w handlu detalicznym .....	9,4	9,5	10,1	6,3	8,7	7,2
Przeciętne wynagrodzenia miesięczne brutto .....	8,2	6,4	8,1	7,3	7,7	6,9
Wskaźniki cen paliw .....	4,3	6,2	3,6	1,6	3,6	2,6
Wskaźniki cen żywności i napojów bezalkoholowych .....	1,9	2,0	1,5	1,4	2,5	1,9

Źródło: opracowanie własne.

Nieco większe znaczenie dla wyjaśnienia stopnia optymizmu konsumentów ma stopa bezrobocia niż liczba ofert pracy. Wynika stąd, że na przewidywania możliwości zaoszczędzenia części dochodu w większym stopniu wpływają informacje z rynku pracy budzące niepokój. Zgodnie z przewidywaniami, znaczny wpływ na ocenę możliwości oszczędzania pieniędzy ma kurs franka szwajcarskiego, od którego zależy wysokość rat kredytów hipotecznych zaciągniętych w tej walucie.

Tak jak można było oczekiwać, na prognozy poziomu bezrobocia najsilniej wpływa sytuacja na rynku pracy. Łączny udział stopy bezrobocia i liczby ofert pracy w  $R^2$  modelu wynosi prawie 20%. Duże znaczenie ma również wartość udzielonych kredytów i pożyczek mieszkaniowych. Branża budowlana jest postrzegana przez respondentów jako siła napędowa gospodarki, a koniunktura w budownictwie mocno wpływa na kondycję pozostałych sektorów i w rezultacie na poziom zatrudnienia.

Wysoka zbieżność występuje w przypadku ocen sytuacji ekonomicznej kraju i sytuacji finansowej gospodarstwa domowego. W obu przypadkach dominujące znaczenie dla respondentów ma kurs franka szwajcarskiego wyjaśniający ponad 14% wariancji każdego ze wskaźników. Wynik ten wiąże się z dynamicznym wzrostem zadłużenia gospodarstw domowych, którego główną część stanowią kredyty mieszkaniowe we frankach szwajcarskich. Wynika stąd, że respondenci są świadomi tego, że wahania kursu tej waluty mogą mieć poważne konsekwencje dla sytuacji finansowej gospodarstwa domowego.

Uporządkowanie według wartości relatywnych wag dla kluczowych czynników wpływających na obie omawiane tu prognozy jest podobne. Kurs euro ma znacznie większy udział w wyjaśnieniu prognozy sytuacji ekonomicznej kraju niż sytuacji finansowej gospodarstwa domowego. Respondenci postrzegają zatem, że kurs euro w większym stopniu oddziałuje na gospodarkę niż na ich własne gospodarstwo domowe.

## Podsumowanie

Otrzymane wyniki wskazują, że opinia konsumentów o poszczególnych wskaźnikach formułuje się pod wpływem szerokiego zbioru danych makroekonomicznych. Oznacza to, że respondenci wyrażając swoją opinię nie kierują się ogólnym odbiorem sytuacji ekonomicznej, ale zwracają uwagę na rozwój sytuacji na różnych rynkach w zależności od tego, jaki wskaźnik poddają ocenie. Potwierdzono hipotezę o tym, że dany czynnik makroekonomiczny ma niejednakowy wpływ na każdy ze wskaźników koniunktury konsumenckiej.

Zastosowanie metody *epsilon* pozwoliło na jednoczesne przebadanie szerokiego zbioru czynników potencjalnie wpływających na kształtowanie się opinii konsumentów. Uwzględnienie korelacji pomiędzy regresorami umożliwiło wyodrębnienie grupy czynników mających decydujący wpływ na opinię respondentów.

Spośród zmiennych opisujących rynek pracy do regresorów tych zaliczono stopę bezrobocia oraz liczbę niewykorzystanych dłużej niż miesiąc ofert pracy. Udział tych zmiennych był szczególnie wysoki w przypadku prognozy bezrobocia, oszczędzania pieniędzy oraz obecnych i planowanych zakupów.

Silny wpływ na kształtowanie się praktycznie wszystkich badanych wskaźników koniunktury ma indeks obrotów w handlu detalicznym. Skala popytu konsumpcyjnego odzwierciedla bowiem zarówno siłę nabywczą ludności, jak i poziom ich optymizmu. Zmniejszenie skali dokonywanych zakupów jest zwykle

związane z rosnącymi obawami dotyczącymi rozwoju sytuacji ekonomicznej, a w szczególności możliwości utraty pracy i zmniejszenia dotychczasowych dochodów.

Wyniki wskazują na niewielką rolę wskaźnika cen konsumpcyjnych *CPI* w wyjaśnieniu stopnia optymizmu ludności. Generalnie, wyższy udział w wyjaśnieniu poszczególnych kategorii miał wskaźnik cen paliw niż *CPI*, pomimo że udział wydatków na paliwa w budżecie jest nieporównywalnie mniejszy niż na żywność i napoje bezalkoholowe. Przyczyn tego zjawiska można upatrywać w prawie dwukrotnie większej wariancji wskaźnika cen paliw niż wskaźnika *CPI*, która świadczy o tym, że ceny paliw podlegają większym zmianom i w rezultacie są obciążone wyższym poziomem niepewności.

---

**mgr Natalia Nehrebecka** — Uniwersytet Warszawski, **mgr Sylwia Grudkowska** — NBP

#### LITERATURA

- Budżety gospodarstw domowych w 2007 r.* (2008), GUS, Warszawa
- Dominitz J., Manski C. (1994), *Using Expectations Data to Study Subjective Income Expectations*, „National Bureau of Economic Research”
- Forsells M., Kenny G. (2002), *The rationality of consumers' inflation expectations: survey based evidence for euro area*, European Central Bank Working Papers, No 163, Frankfurt nad Menem
- Johnson R. (1966), *The minimal transformation to orthonormality*, „Psychometrika”, No 31
- Johnson J. (2000), *A Heuristic Method for Estimating the Relative Weight of Predictor Variables in Multiple Regression*, „Multivariate Behavioral Research”, No 35 (1)
- Koniunktura konsumencka. Tendencje zmian październik 2003—styczeń 2004* (2004), GUS, Warszawa
- Lebreton J., Ployhart R., Ladd R. (2004), *A Monte Carlo Comparison of Relative Importance Methodologies*, „Organizational Research Methods”, No 3, vol. 7
- Maravall A. (2005), *An Application of the automatic procedure of TRAMO and SEATS; Direct versus Indirect Adjustment*, Banco de España Working Papers, No 0524
- Wall M., Rechtsteiner A., Rocha L. (2003), *Singular value decomposition and principal component analysis*, w: Berrar D., Dubitzky W., Granzow M., *A Practical Approach to Microarray Data Analysis*, Springer
- Westerhoff F. (2006), *Nonlinear expectation formation, endogenous business cycles and stylized facts*, „Studies in Nonlinear Dynamics and Econometrics”, vol. 10

#### SUMMARY

*The article presents the relative influence of macroeconomic indicators of consumer confidence in Poland. The analysis was based upon monthly data from the period January 2004—July 2008. The survey was carried out using the epsilon method, which allows calculating the relative weights of indicators, taking into account the correlation between the regressors. The application of this approach has enabled the simultaneous examination of the impact, of each of the*

*13 individual variabibus, on particular consumer activity indicators. The most essential influence on the opinions and forecasts of respondents exents the information on the unemployment rate, the number of job offers, the Swiss franc exchange rate, the retail trade turnover and the value of credits granted to individuals by the banking sector. Minor impact of the credit interests in real terms was noted. The relativcy high amount of average earnings was important only for forecasts of significant purchases and saving money.*

## РЕЗЮМЕ

*В статье представлена оценка относительного влияния макроэкономических факторов на показатели доверия потребителей в Польше. Проанализировались месячные данные за период январь 2004—июль 2008 года. Исчисления проводились с использованием метода epsilon, который позволяет оценить относительные веса показателей с учетом корреляции между независимыми переменными. Применение такого подхода позволило одновременно изучить влияние каждой из 13 переменных на отдельные показатели потребительской конъюнктуры. Решающее влияние на мнение и прогнозы респондентов оказали информации об уровне безработицы, о количестве неиспользованных предложений работы, об обменном курсе швейцарского франка, об оборотах розничной торговли и значении кредитов, предоставленных банками физическим лицам. Незначительную долю имели значения процентов от кредитов в реальном выражении. Относительно большое значение величины средних вознаграждений было отмечено только для прогнозов покупок и экономии денег.*

**Sylwia KOSZNIK-BIERNACKA**

## Modelowanie czasu życia chorych na nowotwory

---

Wykorzystując dwa komplety danych klinicznych dotyczących funkcji przeżycia, wyznaczonej z danych cenzurowanych metodą Kaplana-Meiera (Kaplan, Meier, 1958), w artykule przedstawiono hipotezę opisaną ich komplementarnym rozkładem Makehama (KRM). Rozkład ten został zdefiniowany jako uzupełnienie zmodyfikowanego rozkładu Gomperta. Poprzez takie przekształcenie

otrzymuje się rozkład charakteryzujący się grzbietową funkcją intensywności (początkowo rosnącą, a następnie monotonicznie malejącą), niezbędną, jak pokazują dane medyczne, do modelowania czasu życia chorych na nowotwory (*Metody...*, 2000). Na początku zostaną przedstawione podstawowe cechy przeżywalności dla tego rozkładu (funkcja przeżycia i intensywności oraz uśredniona funkcja intensywności). Inne cechy tego rozkładu można znaleźć w pracy Kosznik-Biernackiej (2006).

Pierwsza z prób ma 80 elementów i dotyczy czasu życia chorych na nowotwory urologiczne, a druga — 7 elementów i dotyczy nowotworów ginekologicznych. W przypadku danych cenzurowanych klasyczne metody estymacji parametrów nie są skuteczne, więc dwa potrzebne parametry rozkładu zostaną oszacowane na podstawie siatki prawdopodobieństwa KRM. Do weryfikacji postawionej hipotezy zostanie zastosowana drobna modyfikacja testu zgodności F. Coxa do porównywania dwóch krzywych przeżycia (Cox, Oakes, 1990). Zamiast porównywać funkcje przeżycia porównamy odpowiednie czasy przeżycia. W opracowaniu zastosowano następujące oznaczenia:  $h(t)$  — funkcja intensywności,  $f(t)$  — funkcja gęstości,  $F(t)$  — dystrybuanta,  $S(t)$  — funkcja przeżycia.

### PODSTAWOWE CHARAKTERYSTYKI PRZEŻYCIA

W badaniach medycznych podstawowymi cechami statystycznymi są:

**Funkcja przeżycia** — określa prawdopodobieństwo, że niepowodzenie (zgon) nastąpi nie wcześniej niż w chwili  $t$ :

$$S(t) = Pr(T \geq t) \quad (1)$$

Dystrybuanta i funkcja przeżycia związane są zależnością:

$$S(t) = 1 - F(t) \quad (2)$$

Dystrybuanta i funkcja przeżycia mogą być również wyrażone przez funkcję gęstości:

$$F(t) = \int_0^t f(u) du \quad S(t) = \int_t^{\infty} f(u) du \quad (3)$$

Funkcja przeżycia  $S(t)$  jest funkcją nierosnącą i spełnia równości  $S(0)=1$ ,  $S(\infty)=0$ .

**Funkcja intensywności:**

$$h(t) = \frac{f(t)}{S(t)} \quad (4)$$

Wartość funkcji intensywności w chwili  $t$  jest prawdopodobieństwem (ryzykiem) niepowodzenia (zgonu) po upływie  $t$ , jeśli do tej chwili ono nie nastąpiło. Prawdziwa jest zależność:

$$h(t) = \frac{-\frac{dS}{dt}}{S(t)} \quad (5)$$

Wyznaczanie funkcji intensywności na podstawie danych doświadczalnych jest często kłopotliwe, ponieważ w jej definicji występuje pochodna (5). Chcąc wyznaczyć pochodną z danych doświadczalnych należy operować skończonym przedziałem  $\Delta t$ . Kłopot polega na tym, że aby możliwie precyzyjnie oddać zmienność funkcji intensywności przedział  $\Delta t$  powinien być wąski, ale jednocześnie liczba zaobserwowanych w tym przedziale niepowodzeń będzie bardzo mała. Jednak liczba ta decyduje o dokładności wyznaczenia interesującego badacza prawdopodobieństwa niepowodzenia w tym przedziale i powinna być ona duża. Opisanych kłopotów można uniknąć, wyznaczając z danych doświadczalnych uśrednioną funkcję intensywności, która zachowuje kształt funkcji intensywności, a przy tym łatwiej ją wyznaczyć.

**Uśredniona funkcja intensywności:**

$$h_{sr}(t) = \frac{1}{t} \int_0^t h(u) du \quad (6)$$

Rozwiązując równanie (5), przy warunku początkowym  $S(0) = 1$ , uzyskuje się związek między uśrednioną funkcją intensywności a funkcją przeżycia lub dystrybuantą:

$$h_{sr}(t) = -\frac{1}{t} \ln(1 - F(t)) = -\frac{1}{t} \ln S(t) \quad (7)$$

Funkcja intensywności (najpierw rosnąca, a potem malejąca) jest często używana do modelowania czasu przeżycia po rozpoznaniu choroby nowotworowej.

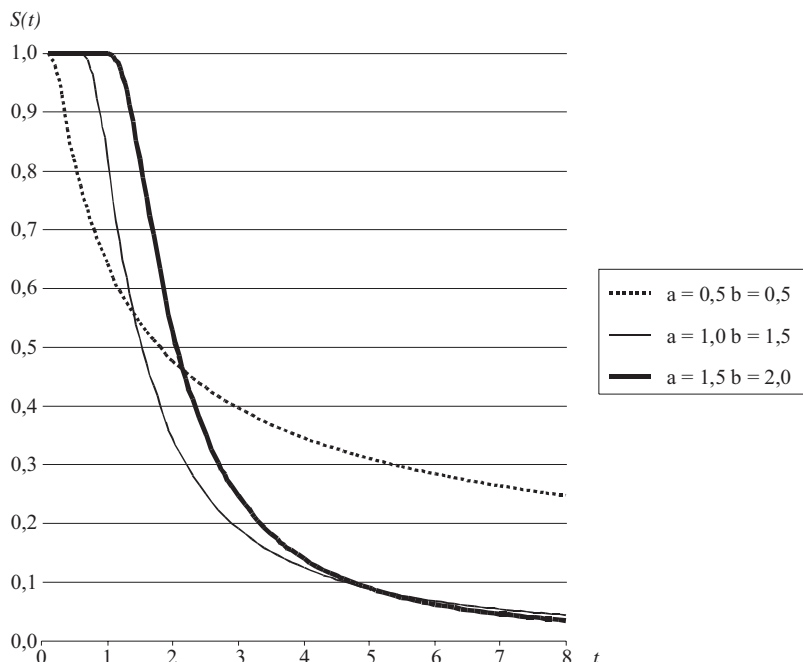
Celem tej publikacji jest zaproponowanie, mało znanego w literaturze, komplementarnego rozkładu Makehama do opisu tego zjawiska. Charakteryzuje się on pożądaną grzbietową funkcją intensywności. Rozkład ten został szczegółowo opisany w pracy Kosznik-Biernackiej (2006).

#### *PODSTAWOWA CHARAKTERYSTYKA PRZEŻYCIA KOMPLEMENTARNEGO ROZKŁADU MAKEHAMA*

Funkcja przeżycia KRM ma postać:

$$S(t) = 1 - \exp\left(1 - \exp\left(\frac{a}{t}\right)^b\right) \quad (8)$$

Wykr. 1. FUNKCJA PRZEŻYCIA KRM DLA WYBRANYCH WARTOŚCI PARAMETRÓW



Źródło: opracowanie własne.

Funkcja gęstości KRM dana jest wzorem:

$$f(t) = \left(\frac{a}{t}\right)^b \cdot \frac{b}{t} \exp\left[\left(\frac{a}{t}\right)^b + 1 - \exp\left(\frac{a}{t}\right)^b\right] \quad (9)$$

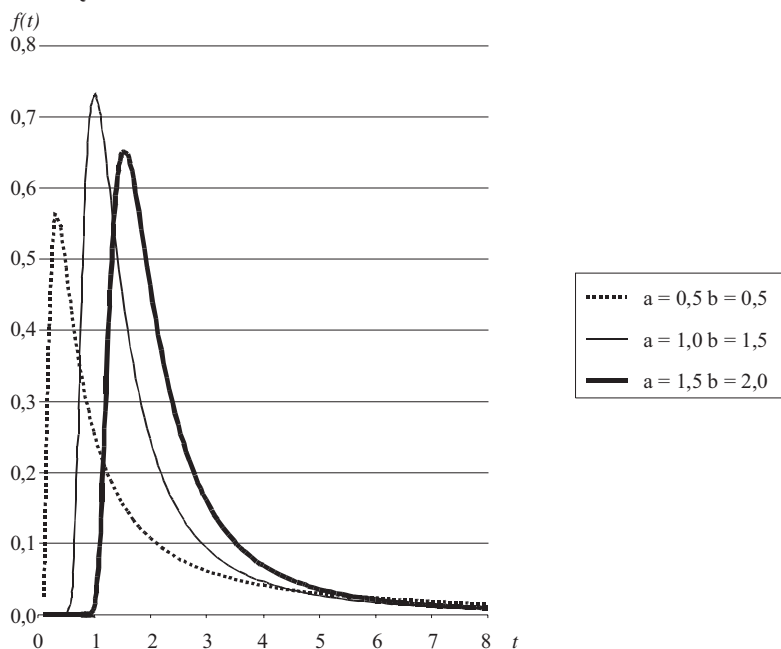
Podstawiając do formuły (4) wzory (8) i (9) otrzymuje się wyrażenie na funkcję intensywności KRM. Charakteryzuje się ona pożądanym do opisu czasu życia chorych na nowotwory przebiegiem grzbietowym czyli najpierw rosnącym, a potem malejącym (wykres 3).

$$h(t) = \frac{\left(\frac{a}{t}\right)^b \cdot \frac{b}{t} \cdot \exp\left[\left(\frac{a}{t}\right)^b + 1 - \exp\left(\frac{a}{t}\right)^b\right]}{1 - \exp\left[1 - \exp\left(\frac{a}{t}\right)^b\right]} \quad (10)$$

Korzystając z (7) uśredniona funkcja ryzyka KRM dana jest wzorem:

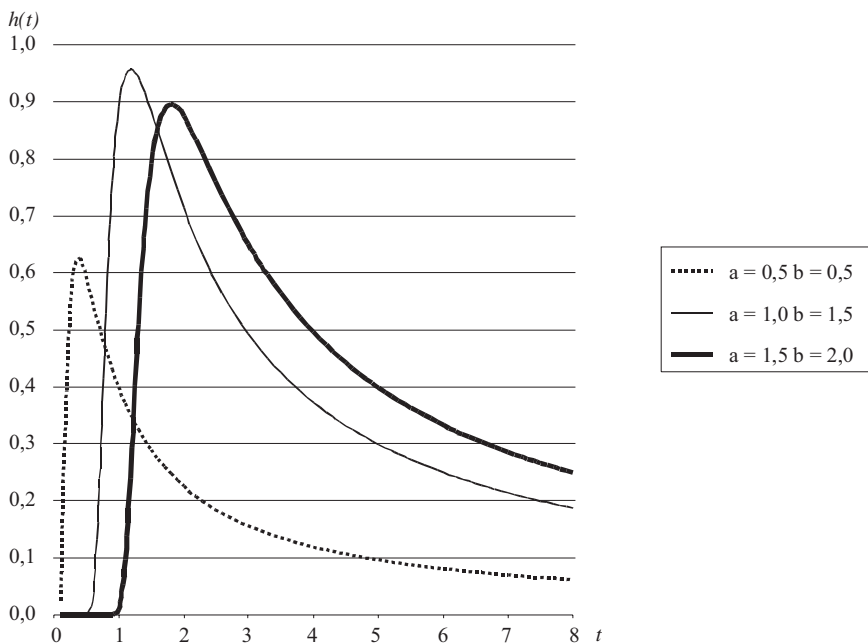
$$h_{sr}(t) = -\frac{1}{t} \cdot \ln\left[1 - \exp\left(1 - \exp\left(\frac{a}{t}\right)^b\right)\right] \quad (11)$$

Wykr. 2. FUNKCJA GĘSTOŚCI KRM DLA WYBRANYCH WARTOŚCI PARAMETRÓW



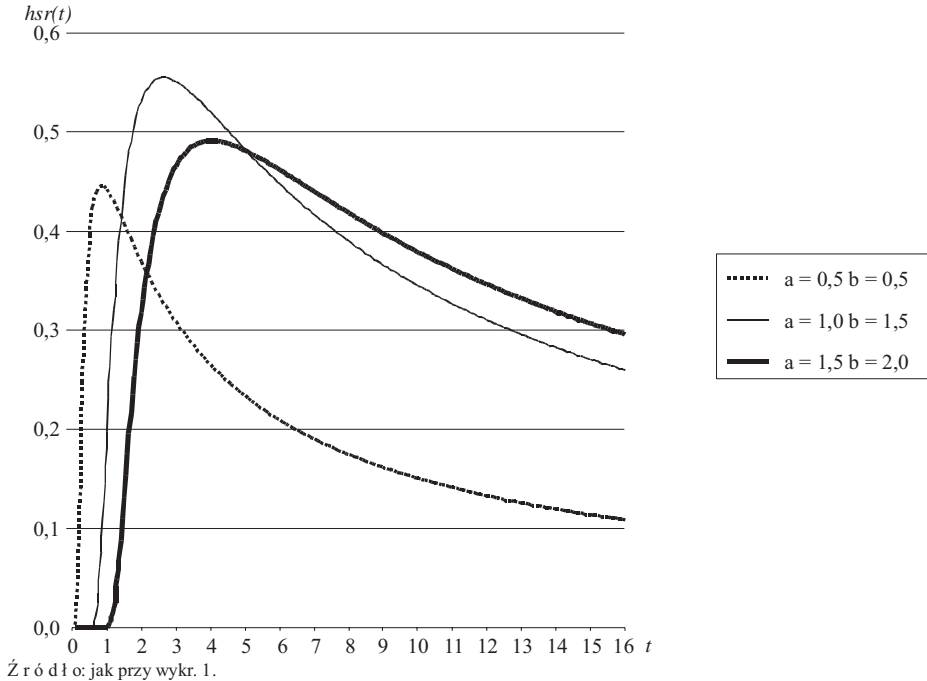
Źródło: jak przy wykr. 1.

Wykr. 3. FUNKCJA INTENSYWNOŚCI KRM DLA WYBRANYCH WARTOŚCI PARAMETRÓW



Źródło: jak przy wykr. 1.

Wykr. 4. UŚREDNIONA FUNKCJA INTENSYWNOŚCI KRM DLA WYBRANYCH WARTOŚCI PARAMETRÓW



Jeśli chodzi o metody estymacji parametrów tego rozkładu, to aby skonstruować siatkę KRM na osi odciętych należy przyjąć transformację:

$$x = \ln t \tag{12}$$

a na osi rzędnych:

$$y(x) = -\ln[\ln(1 - \ln F(x))] \tag{13}$$

Wtedy podstawiając (8) do formuły (13) otrzymuje się:

$$y(x) = -\ln \left[ \ln \left( 1 - \ln \left( \exp \left( 1 - \exp \left( \left( \frac{a}{t} \right)^b \right) \right) \right) \right) \right] = b(\ln t - \ln a) = b(x - \ln a) \tag{14}$$

Właśnie tak dobrane podstawienia (12) i (13) określają siatkę KRM, ponieważ po ich wprowadzeniu do (14) dochodzi się do pożądanego przekształcenia liniowego postaci  $y(x) = bx - d$ .

## PRZYKŁADY

Przedstawiam dwa przykłady oparte na danych klinicznych. Są to wartości funkcji przeżycia w momencie zgonu wyznaczone z danych cenzurowanych metodą Kaplana-Meiera (Balicki, 2006; Frątczak i in., 2005).

### Przykład 1

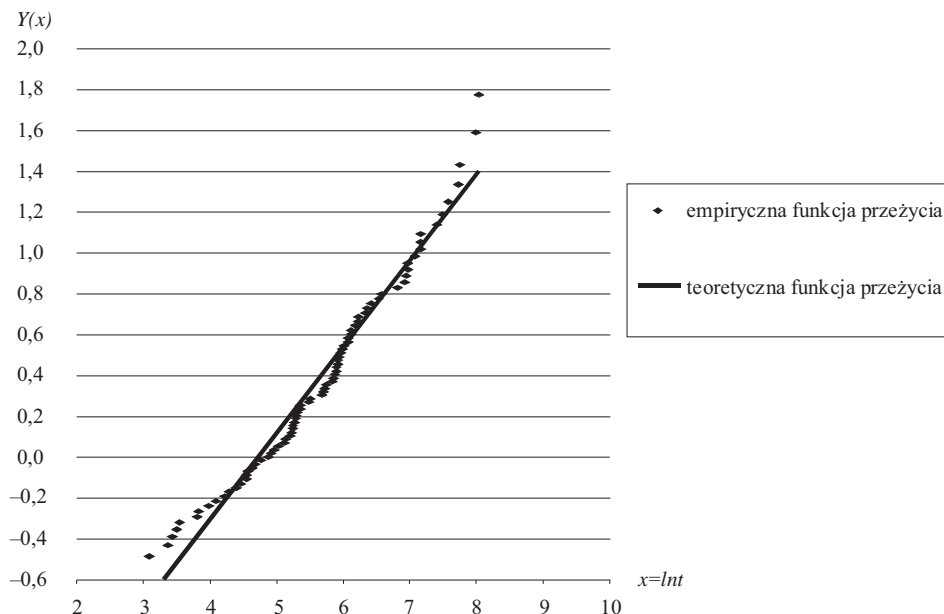
**TABL. 1. DANE KLINICZNE DOTYCZĄCE CZASU ŻYCIA  
CHORYCH NA NOWOTWORY UROLOGICZNE**

Lp.	Czas życia (dni)	Funkcja przeżycia $S(t)$	Lp.	Czas życia (dni)	Funkcja przeżycia $S(t)$
1 .....	22	0,982759	41 .....	314	0,637931
2 .....	29	0,974138	42 .....	341	0,629310
3 .....	31	0,965517	43 .....	349	0,620690
4 .....	33	0,956897	44 .....	359	0,612069
5 .....	35	0,948276	45 .....	365	0,603448
6 .....	45	0,939655	46 .....	367	0,594828
7 .....	46	0,931034	47 .....	373	0,586207
8 .....	54	0,922414	48 .....	377	0,577586
9 .....	60	0,913793	49 .....	380	0,568966
10 .....	68	0,905172	50 .....	394	0,560345
11 .....	73	0,896552	51 .....	400	0,551724
12 .....	81	0,887931	52 .....	412	0,543103
13 .....	87	0,879310	53 .....	436	0,534344
14 .....	95	0,870690	54 .....	436	0,525584
15 .....	95	0,862069	55 .....	455	0,516522
16 .....	96	0,853448	56 .....	458	0,507460
17 .....	103	0,844828	57 .....	489	0,498234
18 .....	109	0,836207	58 .....	502	0,489007
19 .....	117	0,827586	59 .....	512	0,479781
20 .....	130	0,818966	60 .....	564	0,470373
21 .....	138	0,810345	61 .....	579	0,460966
22 .....	144	0,801724	62 .....	614	0,451558
23 .....	154	0,793103	63 .....	690	0,441951
24 .....	166	0,784483	64 .....	715	0,432130
25 .....	172	0,775862	65 .....	910	0,421326
26 .....	182	0,767241	66 .....	1022	0,409939
27 .....	186	0,758621	67 .....	1049	0,398552
28 .....	189	0,750000	68 .....	1065	0,387165
29 .....	190	0,741379	69 .....	1070	0,375778
30 .....	193	0,732759	70 .....	1184	0,364390
31 .....	198	0,724138	71 .....	1284	0,352636
32 .....	199	0,715517	72 .....	1290	0,340476
33 .....	203	0,706897	73 .....	1298	0,328316
34 .....	210	0,698276	74 .....	1647	0,314042
35 .....	212	0,689655	75 .....	1812	0,299087
36 .....	240	0,681034	76 .....	1939	0,281494
37 .....	249	0,672414	77 .....	2262	0,259840
38 .....	294	0,663793	78 .....	2332	0,236219
39 .....	299	0,655172	79 .....	2977	0,202473
40 .....	309	0,646552	80 .....	3100	0,168728

Źródło: opracowanie własne.

KRM charakteryzuje się pożądanym do opisu czasu życia chorych na chorobę nowotworową grzbietowym przebiegiem funkcji intensywności, zatem postawiono hipotezę, że dane kliniczne z tabl. 1 można estymować KRM z wyznaczonymi wartościami parametrów. Dane te są danymi cenzurowanymi i nie znamy licznosci badanej próby, więc parametry rozkładu będą estymowane z siatki KRM.

**Wykr. 5. DANE KLINICZNE Z TABL1 NA SIATCE KRM**



Źródło: jak przy wykr. 1.

Przeprowadzając klasyczną analizę regresji dopasowano do danych klinicznych na siatce model liniowy.

**ZESTAWIENIE (1) WYNIKÓW PRZEPROWADZONEJ ANALIZY REGRESJI  
(z przykładu 1)**

**Statystyka regresji**

Współczynnik korelacji wielorakiej (wielokrotność $R$ ) .....	0,985646137
Współczynnik determinacji ( $R^2$ ) .....	0,971498308
Skorygowany współczynnik determinacji (dopasowany $R^2$ ) .....	0,971132901
Błąd standardowy .....	0,085004131
Obserwacje .....	80

**ZESTAWIENIE (1) WYNIKÓW PRZEPROWADZONEJ ANALIZY REGRESJI (dok.)**

**Analiza wariancji**

Wyszczególnienie	Stopnie swobody <i>df</i>	Suma kwadratów <i>SS</i>	Średnia kwadratów <i>MS</i>	Statystyka <i>F</i>	Istotność <i>F</i>
Regresja .....	1	19,21082747	19,21083	2658,6796	5,01852E-62
Resztkowy .....	78	0,563604785	0,007226	x	x
Razem .....	79	19,77443226	x	x	x

**Oszacowania parametrów modelu i ich statystyczne charakterystyki**

Parametry modeli	Współczynniki modelu	Błąd standardowy	Statystyka <i>t</i> -Studenta	Graniczny poziom istotności <i>p</i> dla $\alpha = 0,05$	Dolna granica przedziału ufności dla $\alpha = 0,05$	Górna granica przedziału ufności dla $\alpha = 0,05$
Przecięcie ...	-1,991311125	0,047223702	-42,1676	1,904E-55	-2,085326292	-1,897295958
$x = \ln t$ .....	0,422019527	0,008184639	51,56239	5,019E-62	0,405725162	0,438313892

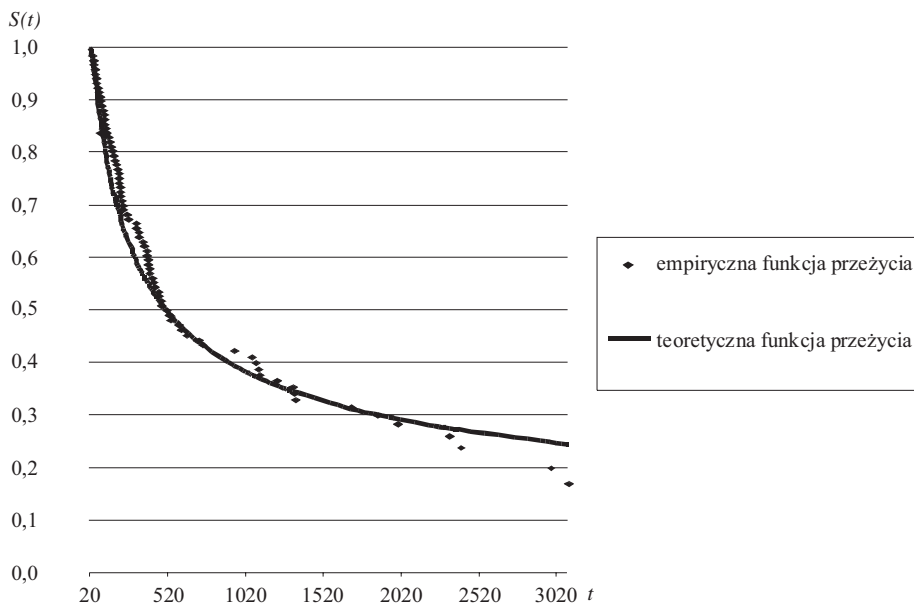
U w a g a. Analiza regresji została przeprowadzona przy pomocy dodatku Excela *Analiza danych* → *Regresja*.

Wszystkie estymowane wartości stałe w modelu są istotne statystycznie, co daje ostatecznie model postaci:

$$y = 0,422019527x - 1,991311125 \quad (15)$$

Model ten tłumaczy ponad 97% zmienności skumulowanej proporcji przeżycia, przez co można go uznać za model wysokiej jakości. Zatem korzystając z własności siatki prawdopodobieństwa rozkładu otrzymuje się oszacowania parametrów KRM. Otrzymane wartości, to:  $a = 111,9452$ ,  $b = 0,422$ .

**Wykr. 6. EMPIRYCZNA I TEORETYCZNA FUNKCJA PRZEŻYCIA DLA ROZPATRYWANYCH DANYCH KLINICZNYCH (w przykładzie 1)**



Źródło: jak przy wykr. 1.

Niestety nie można postawionej hipotezy zweryfikować typowym testem zgodności, ponieważ empiryczne funkcje przeżycia zostały wyznaczone metodą Kaplana-Meiera z danych cenzurowanych i dane są tylko wartości funkcji przeżycia w momencie zgonu.

Zatem weryfikacja hipotezy zerowej zostanie przeprowadzona testem F. Coxa (Stanisz, 2007) do porównywania dwóch krzywych przeżycia. Oszacowane wartości parametrów KRM, to:  $a = 111,9452$ ,  $b = 0,422$ . Załóżmy, że prawdziwa jest równość:

$$S(t_i) = S_{teor}(t_i) \quad (16)$$

gdzie  $S_{teor}(t)$  dana jest wzorem (8) z parametrami  $a = 111,9452$ ,  $b = 0,422$ . Jeżeli założenie (16) jest prawdziwe, to z matematycznego punktu widzenia odpowiednie funkcje odwrotne też są równe:

$$S^{-1}(t_i) = S^{-1}_{teor}(t_i) \quad (17)$$

A skoro tak jest, to można wyznaczyć odpowiedni czas przeżycia, zakładając że empiryczna funkcja przeżycia ma KRM z parametrami  $a = 111,9452$ ,  $b = 0,422$ . Po wykonaniu stosownych obliczeń zostały one przedstawione w tabl. 2.

**TABL. 2. TEORETYCZNY CZAS PRZEŻYCIA DLA ROZPATRYWANYCH  
(w przykładzie 1) DANYCH KLINICZNYCH**

Lp.	$T_i$	Lp.	$T_i$	Lp.	$T_i$	Lp.	$T_i$
1 .....	35,61	21 .....	117,64	41 .....	259,71	61 .....	633,64
2 .....	40,37	22 .....	122,57	42 .....	270,32	62 .....	668,49
3 .....	44,63	23 .....	127,67	43 .....	281,41	63 .....	706,71
4 .....	48,63	24 .....	132,92	44 .....	293,04	64 .....	748,79
5 .....	52,48	25 .....	138,36	45 .....	305,22	65 .....	798,97
6 .....	56,25	26 .....	143,98	46 .....	317,99	66 .....	856,78
7 .....	59,97	27 .....	149,80	47 .....	331,40	67 .....	920,29
8 .....	63,69	28 .....	155,82	48 .....	345,49	68 .....	990,23
9 .....	67,41	29 .....	162,07	49 .....	360,30	69 .....	1067,48
10 .....	71,17	30 .....	168,55	50 .....	375,88	70 .....	1153,06
11 .....	74,96	31 .....	175,28	51 .....	392,28	71 .....	1251,42
12 .....	78,82	32 .....	182,27	52 .....	409,57	72 .....	1365,53
13 .....	82,74	33 .....	189,53	53 .....	428,11	73 .....	1494,28
14 .....	86,74	34 .....	197,08	54 .....	447,70	74 .....	1667,55
15 .....	90,82	35 .....	204,95	55 .....	469,16	75 .....	1880,04
16 .....	95,00	36 .....	213,14	56 .....	491,92	76 .....	2180,83
17 .....	99,29	37 .....	221,67	57 .....	516,54	77 .....	2650,51
18 .....	103,68	38 .....	230,58	58 .....	542,76	78 .....	3339,72
19 .....	108,20	39 .....	239,87	59 .....	570,69	79 .....	4843,88
20 .....	112,85	40 .....	249,57	60 .....	601,10	80 .....	7502,06

Źródło: opracowanie własne.

Przeprowadzając odpowiednie obliczenia w programie Statistica testu F. Coxa otrzymuje się następujące wyniki:  $T_1 = 78,92560$ ;  $T_2 = 81,07440$ ;  $F(160, 160) = 1,027226$ ;  $p = 0,43265$ .

Zatem na poziomie istotności  $\alpha = 0,05$  nie mamy podstaw do odrzucenia hipotezy zerowej.

### Przykład 2

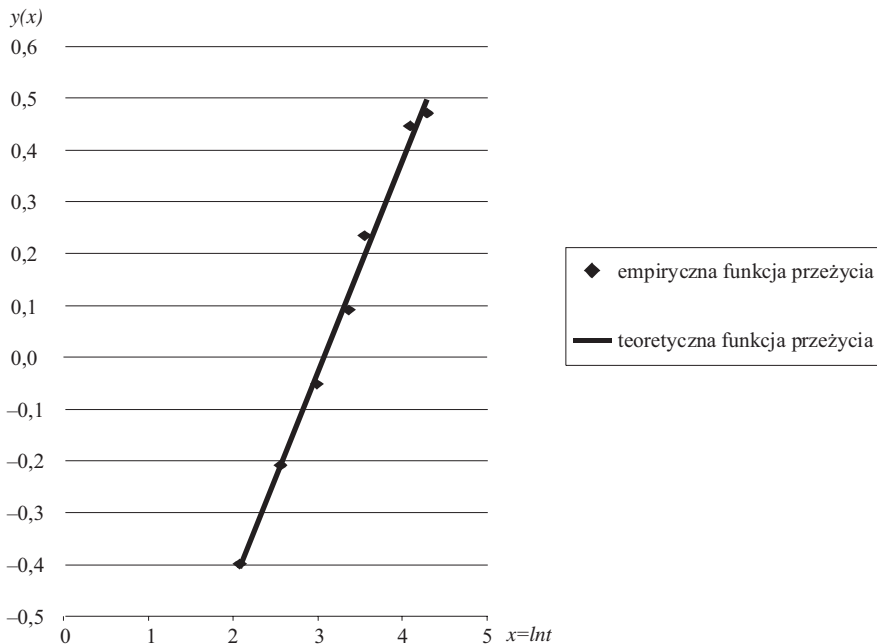
**TABL. 3. DANE KLINICZNE DOTYCZĄCE CZASU ŻYCIA CHORYCH NA NOWOTWORY GINEKOLOGICZNE**

Lp.	Czas życia (miesiące)	Funkcja przeżycia
1 .....	8	0,968
2 .....	13	0,911
3 .....	20	0,846
4 .....	29	0,775
5 .....	35	0,700
6 .....	60	0,592
7 .....	73	0,580

Źródło: opracowanie własne.

Postawiono hipotezę, że dane kliniczne z tabl. 3 też można estymować KRM z wyznaczonymi wartościami parametrów. Podobnie jak w przykładzie 1, dane te są danymi cenzurowanymi i nie znamy licznosci badanej próby, więc parametry rozkładu będą estymowane na podstawie siatki KRM.

**Wykr. 7. DANE KLINICZNE Z TABL. 4 NA SIATCE KRM**



Źródło: jak przy wyk. 1.

Przeprowadzając klasyczną analizę regresji dopasowano do danych empirycznych na siatce model liniowy.

**ZESTAWIENIE (2) WYNIKÓW PRZEPROWADZONEJ ANALIZY REGRESJI  
(z przykładu 2)**

**Statystyka regresji**

Współczynnik korelacji wielorakiej (wielokrotność $R$ ) .....	0,99653536
Współczynnik determinacji ( $R^2$ ) .....	0,99308272
Skorygowany współczynnik determinacji (dopasowany $R^2$ ) .....	0,99169926
Błąd standardowy .....	0,02978963
Obserwacje .....	7

**Analiza wariancji**

Wyszczególnienie	Stopnie swobody $df$	Suma kwadratów $SS$	Średnia kwadratów $MS$	Statystyka $F$	Istotność $F$
Regresja .....	1	0,637015655	0,6370157	717,827	1,35455E-06
Resztkowy .....	5	0,004437111	0,0008874	x	x
Razem .....	6	0,641452767	x	x	x

**Oszacowania parametrów modelu i ich statystyczne charakterystyki**

Parametry modeli	Współczynniki modelu	Błąd standardowy	Statystyka $t$ -Studenta	Graniczny poziom istotności $p$ dla $\alpha = 0,05$	Dolna granica przedziału ufności dla $\alpha = 0,05$	Górna granica przedziału ufności dla $\alpha = 0,05$
Przecięcie ...	-1,25843554	0,051340011	-24,51179	2,11E-06	-1,390409242	-1,12646184
$x = \ln t$ .....	0,40937702	0,015279656	26,792294	1,35E-06	0,370099412	0,448654622

U w a g a. Jak przy zestawieniu (1).

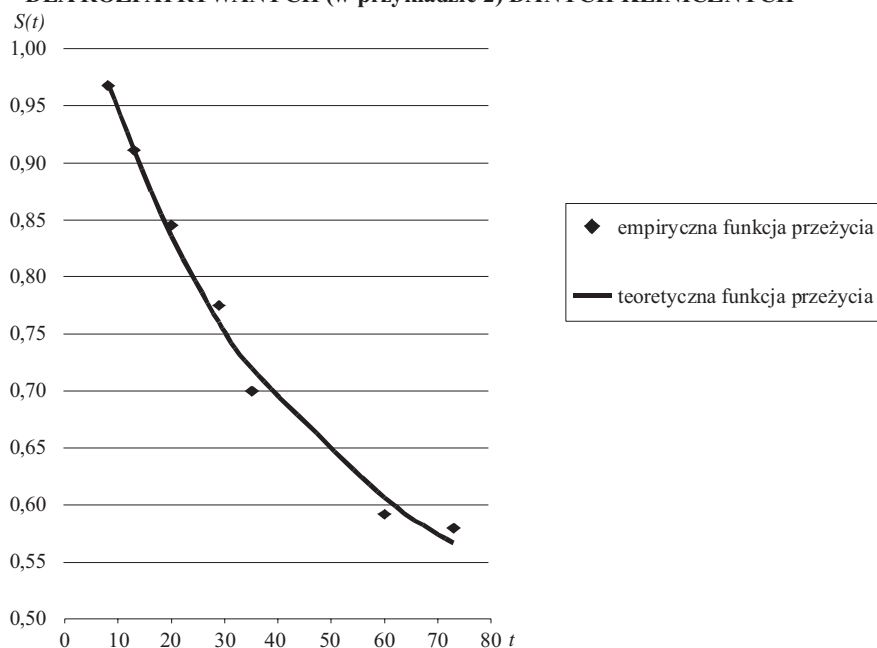
Wszystkie estymowane wartości stałe w modelu są istotne statystycznie, co daje ostatecznie model postaci:

$$y = 0,40937702 x - 1,25843554 \quad (18)$$

Model ten tłumaczy ponad 99% zmienności skumulowanej proporcji przeżycia, przez co można go uznać za model bardzo wysokiej jakości. Zatem korzystając z własności siatki prawdopodobieństwa rozkładu otrzymuje się oszacowania parametrów KRM. Otrzymane wartości, to:  $a = 21,62319$ ,  $b = 0,4094$ .

Podobnie jak w przykładzie 1, hipoteza zerowa o prawdziwości równości (16), gdzie  $S_{teor}(t)$  dana jest wzorem (8) z parametrami  $a = 21,62319$ ,  $b = 0,4094$ , zostanie zweryfikowana testem F. Coxa. Odpowiedni czas przeżycia rozpatrywanych przykładów, zakładając że empiryczna funkcja przeżycia ma KRM z parametrami  $a = 21,62319$ ,  $b = 0,4094$ , przedstawiono w tabl. 4.

**Wykr. 8. EMPIRYCZNA I TEORETYCZNA FUNKCJA PRZEŻYCIA  
DLA ROZPATRYWANYCH (w przykładzie 2) DANYCH KLINICZNYCH**



Źródło: jak przy wykr. 1.

**TABL. 4. TEORETYCZNY CZAS PRZEŻYCIA DLA ROZPATRYWANYCH  
(w przykładzie 2) DANYCH KLINICZNYCH**

Lp.	$T_i$
1 .....	8,158777
2 .....	13,01779
3 .....	19,03148
4 .....	26,99024
5 .....	38,44246
6 .....	64,46835
7 .....	68,26225

Źródło: opracowanie własne.

Przeprowadzając stosowne obliczenia w środowisku programistycznym Statistica, otrzymano:  $T_1 = 7,276402$ ;  $T_2 = 6,723598$ ;  $F(14, 14) = 1,082218$ ;  $p = 0,44228$ .

Zatem na poziomie istotności  $\alpha = 0,05$  nie ma podstaw do odrzucenia hipotezy zerowej.

## Wnioski

Przedstawiony mało znany KRM bardzo dobrze opisuje dane kliniczne zarówno dla dużej, jak i małej próby. Rozkład ten może więc być rekomendowany do modelowania czasu życia chorych na nowotwory.

dr Sylwia Kosznik-Biernacka — Akademia Pomorska w Słupsku

## LITERATURA

- Balicki A. (2006), *Analiza przeżycia i tablice wymieralności*, PWE, Warszawa
- Cox D., Oakes D. (1990), *Analysis of survival date*, Chapman and Hall, New York
- Frątczak E., Gach-Ciepiela U., Babiker H. (2005), *Analiza historii zdarzeń*, SGH, Warszawa
- Kaplan E. L., Meier P. (1958), *Nonparametric estimation from incomplete observations*, „Journal of the American Statistical Association”
- Kosznik-Biernacka S. (2006), *Rodzina rozkładów Makehama*, PAP, Słupsk
- Metody oceny i porządkowania ryzyka w ubezpieczeniach życiowych* (2000), AE Wrocław, pod red. S. Stasiewicz
- Stanisz A. (2007), *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny*, tom 3, „Analizy wielowymiarowe”, StatSoft, Kraków

## SUMMARY

*The article presents a medical data analysis and a model of the life expectancy for cancer patients. The model function is of the initial rising and than monotonic declining distribution. The Makeham's Complementary Distribution characterizes such function intensity. The base characteristic of the life expectancy by the Makeham's Complementary Distribution as well as two examples for big and small samples of using this distribution for describing the clinical data are presented in the article.*

## РЕЗЮМЕ

*В статье представляется анализ медицинских данных, из которых вытекает, что для моделирования времени жизни страдающих от рака заболеваний следует применить распределение с растущей в начале функцией интенсивности, а затем однообразно (монотонно) уменьшающейся. Именно таким, желательным для описания этого явления, ходом функции интенсивности характеризуется комплементарное распределение Мейкхама (KRM). В статье представляется основная характеристика прожития комплементарного распределения Мейкхама а также два примера как большой, так и ограниченной выборки применения этого распределения для описания клинических данных.*