POLSKA AKADEMIA NAUK
KOMITET STATYSTYKI I EKONOMETRII

PRZEGLĄD STATYSTYCZNY

STATISTICAL REVIEW

TOM 63

4

2016

WARSZAWA 2016

WYDAWCA

Komitet Statystyki i Ekonometrii Polskiej Akademii Nauk

Wydanie publikacji dofinansowane przez Ministra Nauki i Szkolnictwa Wyższego

Strona WWW „Przegląd Statystyczny”:
http://www.przegladstatystyczny.pan.pl

Nakład 140 egz.

DOM WYDAWNICZY ELIPSA

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

Druk:
Wrocławska Drukarnia Naukowa PAN Sp. z o.o.
im. Stanisława Kulczyńskiego
53-505 Wrocław, ul. Lelewela 4

SPIS TREŚCI

CONTENTS

DOI: 10.5604/01.3001.0014.1213

PRZEGLĄD STATYSTYCZNY
R. LXIII – ZESZYT 4 – 2016

EMIL PANEK[1]

GOSPODARKA GALE'A Z WIELOMA MAGISTRALAMI. „SŁABY" EFEKT MAGISTRALI

1. WSTĘP

W teorii magistral dowodzi się, że bez względu na stan wyjściowy gospodarki optymalne procesy wzrostu są w długich okresach czasu zawsze zbieżne do pewnej wzorcowej ścieżki zwanej magistralą (produkcyjną, kapitałową, konsumpcyjną), na której gospodarka osiąga najwyższe tempo równomiernego wzrostu[2]. W stacjonarnych modelach dynamiki ekonomicznej typu Neumanna-Gale'a-Leontiefa magistrala jest zazwyczaj określona jednoznacznie, a jej obrazem geometrycznym w przestrzeni stanów gospodarki jest półprosta zwana promieniem von Neumanna. Pisaliśmy o tym m.in. w artykułach Panek (2015a, 2016) sugerując, że założenie jednoznaczności magistrali można osłabić, ale niestety kosztem wzrostu złożoności modelu. Przykład takiego uogólnienia prezentujemy poniżej. Co ciekawe, złożoność modelu szczególnie przy tym nie rośnie.

Zajmiemy się n – produktową gospodarką typu Gale'a, w której pojedynczą magistralę zastąpimy wiązką magistral, nazywaną dalej umownie magistralą wielopasmową. Ustalimy warunki istnienia magistrali wielopasmowej w gospodarce Gale'a oraz jej niektóre własności, ale przede wszystkim pokażemy, że podobnie jak w klasycznych modelach z pojedynczą magistralą, w gospodarce Gale'a z magistralą wielopasmową obserwujemy specyficzną stabilność optymalnych procesów wzrostu, nazywaną w literaturze „słabym" efektem magistrali.

Prezentowany model nawiązuje do wersji przedstawionej w książce Panek (2003, rozdz. 5, punkt 5.1).

---

[1] Uniwersytet Ekonomiczny w Poznaniu, Wydział Informatyki i Gospodarki Elektronicznej, Katedra Ekonomii Matematycznej, al. Niepodległości 10, 60-967 Poznań, Polska, e-mail: emil.panek@ue.poznan.pl.

[2] Zob. np. McKenzie (1976; 1998; 2005, rozdz. 26), Nikaido (1968, rozdz. IV), Panek (2003, rozdz. 5, 6), Takayama (1985, rodz. 7).

2. PRZESTRZEŃ PRODKCYJNA GALE'A. DEFINICJA, WYBRANE WŁASNOŚCI

Standardowa n – produktowa przestrzeń produkcyjna Gale'a Z (inaczej: zbiór technologiczny) jest stożkiem wypukłym zawartym w R_+^{2n}, z wierzchołkiem w 0. Jego elementami są $2n$-wymiarowe wektory $(x, y) = (x_1, \dots, x_n, y_1, \dots, y_n) \geqq 0$[3], o których mówimy, że opisują dopuszczalne procesy produkcji w gospodarce Gale'a; x – nazywamy wektorem zużycia (nakładów), y – wektorem produkcji (wyników). Zapis $(x, y) \in Z$ oznacza, że w gospodarce, której technologię opisuje przestrzeń produkcyjna Z, z wektora nakładów x możliwe jest wytworzenie wektora produkcji y.

Procesy produkcji podporządkowane są następującym regułom:

(G1) $\forall (x, y) \in Z \; \forall \lambda \geq 0 \; (\lambda(x, y) \in Z)$

(warunek proporcjonalności nakładów i wyników),

(G2) $\forall \left(x^i, y^i\right) \in Z, i = 1{,}2 \left((x^1 + x^2, y^1 + y^2) \in Z\right)$

(warunek addytywności procesów produkcyjnych),

(G3) $\forall (x, y) \in Z \; (x = 0 \Longrightarrow y = 0)$

(warunek „braku rogu obfitości"),

(G4) $\forall (x, y) \in Z \; (x' \geqq x \Longrightarrow (x', y) \in Z)$

(możliwość marnotrawstwa nakładów),

(G5) $\forall (x, y) \in Z \; (y' \leqq y \Longrightarrow (x, y') \in Z)$

(możliwość marnotrawstwa mocy produkcyjnych),

(G6) $\forall \left(x^i, y^i\right) \in Z, i = 1{,}2, \dots \left(\left(x^i, y^i\right) \underset{i}{\rightarrow} (\bar{x}, \bar{y}) \Longrightarrow (\bar{x}, \bar{y}) \in Z\right)$

(domkniętość przestrzeni produkcyjnych).

Zauważmy, że jeżeli $(x, y) \in Z$ oraz $(x, y) \neq 0$, to $x \neq 0$. Dalej interesują nas nietrywialne (niezerowe) procesy $(x, y) \in Z \setminus \{0\}$.

Liczbę

$$\alpha(x, y) = \max\{\alpha | \alpha x \leqq y\}$$

[3] Jeżeli $a, b \in R^n$, to zapis $a \geqq b$ oznacza, że dla każdego i zachodzi: $a_i \geq b_i$. Pisząc $a \geq b$ rozumiemy, że $a \geqq b$ oraz $a \neq b$.

nazywamy wskaźnikiem technologicznej efektywności procesu $(x, y) \in Z \setminus \{0\}$. Funkcja α jest ciągła i dodatnio jednorodna stopnia 0 na $Z \setminus \{0\}$. Stąd i z twierdzenia Weierstrassa o istnieniu maksimum funkcji ciągłej na zbiorze zwartym otrzymujemy wniosek, że istnieje rozwiązanie $(\bar{x}, \bar{y}) \in Z \setminus \{0\}$ zadania

$$\max_{(x,y)\in Z\setminus\{0\}} \alpha(x, y) = \max_{(x,y)\in\Omega(1)} \alpha(x, y) = \alpha(\bar{x}, \bar{y}) = \alpha_M, \tag{1}$$

gdzie $\Omega(1) = \{(x, y) \in Z | \, \|(x, y)\| = 1\}$ (tutaj i dalej $\|a\| = \sum_i |a_i|$).[4] Wobec dodatniej jednorodności stopnia 0 funkcji α rozwiązanie to jest określone z dokładnością do mnożenia przez dowolną stałą dodatnią (z dokładnością do struktury):

jeżeli

$$\alpha(\bar{x}.\bar{y}) = \max_{(x,y)\in Z\setminus\{0\}} \alpha(x, y) = \alpha_M,$$

to także $\forall \lambda > 0$

$$\alpha(\lambda\bar{x}.\lambda\bar{y}) = \max_{(x,y)\in Z\setminus\{0\}} \alpha(x, y) = \alpha_M. \tag{2}$$

Proces $(\bar{x}, \bar{y})$ będący rozwiązaniem zadania (1) nazywamy optymalnym procesem produkcji w gospodarce Gale'a. Liczbę α_M nazywamy optymalnym wskaźnikiem technologicznej efektywności produkcji.

Oznaczmy przez $Z_{opt} \subset Z \setminus \{0\}$ zbiór wszystkich optymalnych procesów produkcji:

$$Z_{opt} = \{(\bar{x}, \bar{y}) \in Z \setminus \{0\} | \, \alpha(\bar{x}, \bar{y}) = \alpha_M\}.$$

O jego kształcie mówi poniższe twierdzenie.

□ **Twierdzenie 1.** Przy założeniach **(G1)**–**(G6)** zbiór Z_{opt} jest stożkiem wypukłym nie zawierającym 0.

Dowód. Weźmy dowolną parę optymalnych procesów $(\bar{x}^i, \bar{y}^i) \in Z_{opt}$, $i = 1, 2$, oraz takie liczby $\lambda_1, \lambda_2 \geq 0$, że $\lambda_1 + \lambda_2 > 0$. Niech $(\bar{x}, \bar{y}) = \lambda_1(\bar{x}^1, \bar{y}^1) + \lambda_2(\bar{x}^2, \bar{y}^2)$. Wówczas $(\bar{x}, \bar{y}) \in Z \setminus \{0\}$, czyli

$$\alpha(\bar{x}.\bar{y}) \leq \alpha_M. \tag{3}$$

[4] Zob. np. Panek (2003, tw. 5.2). Zbiór $\Omega(1)$ jest zwarty. W twierdzeniu 5.2 zamiast niego mamy (też zwarty) zbiór $Q = \{(x, y) \in Z | \, \|(x, y)\| = 1\}$).

Z założenia $\alpha(\bar{x}^1, \bar{y}^1) = \alpha(\bar{x}^2, \bar{y}^2) = \alpha_M$, więc

$$\alpha_M \bar{x}^1 \leqq \bar{y}^1 \text{ oraz } \alpha_M \bar{x}^2 \leqq \bar{y}^2$$

i wobec tego

$$\alpha_M \bar{x} = \alpha_M(\lambda_1 \bar{x}^1 + \lambda_2 \bar{x}^2) \leqq \lambda_1 \bar{y}^1 + \lambda_2 \bar{y}^2 = \bar{y},$$

skąd wnioskujemy, że:

$$\alpha(\bar{x}.\bar{y}) \geq \alpha_M. \tag{4}$$

Z (3), (4) otrzymujemy $\alpha(\bar{x}.\bar{y}) = \alpha_M$, zatem $(\bar{x}, \bar{y}) \in Z_{opt}$. ■

Jeżeli $(\bar{x}, \bar{y}) \in Z_{opt}$, to wobec **(G4)**, **(G5)** również $(\bar{x}, \alpha_M \bar{x}) \in Z_{opt}$ oraz $(\bar{y}, \alpha_M \bar{y}) \in Z_{opt}$. Jednakowoż, bez dodatkowych warunków, o optymalnym wskaźniku technologicznej efektywności α_M możemy jedynie powiedzieć, że jest nieujemny. Aby wykluczyć ten nierealistyczny przypadek zakładamy, że spełniony jest następujący postulat, który nazywamy warunkiem silnej regularności[5]:

(G7) $\forall(\bar{x}, \bar{y}) \in Z_{opt}\ (\bar{y} > 0)$

głoszący, że w optymalnych procesach produkcji wytwarzane są wszystkie towary. Oczywiście, wówczas $\alpha_M > 0$. Ponadto, z **(G7)**, **(G4)** wynika, że istnieją takie procesy $(\bar{x}, \bar{y}) > 0$, w których produkcja $\bar{y}$ jest (po wszystkich współrzędnych) α_M – wielokrotnością nakładów $\bar{x}$:

$$\exists(\bar{x}, \bar{y}) \in Z_{opt}(\bar{y} = \alpha_M \bar{x} > 0). \tag{5}$$

Weźmy dowolny proces produkcji $(x, y) \in Z \setminus \{0\}$ z wektorem $y \neq 0$. O wektorze

$$s = \frac{y}{\|y\|} = \left(\frac{y_1}{\|y\|}, \frac{y_2}{\|y\|}, \ldots, \frac{y_n}{\|y\|}\right)$$

mówimy, że charakteryzuje strukturę produkcji w procesie (x,y).

Oznaczamy przez S następujący zbiór wektorów struktury produkcji we wszystkich optymalnych procesach $(\bar{x}, \bar{y}) \in Z_{opt}$:

$$S = \left\{s \mid \exists(\bar{x}, \bar{y}) \in Z_{opt}\left(s = \frac{\bar{y}}{\|\bar{y}\|}\right)\right\}. \tag{6}$$

W standardowym modelu Gale'a zbiór ten redukuje się do punktu. O jego (niektórych) własnościach mówi kolejne twierdzenie.

[5] Słabszą wersję tego warunku sformułujemy w punkcie 6.

□ **Twierdzenie 2.** Przy przyjętych założeniach zbiór *S*:
(i) jest niepusty, zwarty i wypukły,
(ii) zawiera wyłącznie wektory dodatnie.

Dowód. (i) Zbiór *S* jest niepusty, gdyż istnieje rozwiązanie zadania (1).

(Zwartość) Weźmy dowolny ciąg wektorów $s^i \in S$, $i = 1, 2, \ldots, \infty$. Ponieważ $S \subset S_+^n(1) = \{s \in R_+^n | \|s\| = 1\}$ i simpleks $S_+^n(1)$ jest zbiorem zwartym, więc

$$\exists \{s^{i_j}\}_{j=1}^{\infty} \left(s^{i_j} \underset{j}{\to} \bar{s} \in S_+^n(1) \right).$$

Niech $\eta^{i_j} = \alpha_M s^{i_j}$. Wówczas

$$\left(s^{i_j}, \eta^{i_j}\right) = \left(s^{i_j}, \alpha_M s^{i_j}\right) \in Z \setminus \{0\}, j = 1,2, \ldots, \infty,$$

$$\left(s^{i_j}, \eta^{i_j}\right) \underset{j}{\to} (\bar{s}, \bar{\eta}) = (\bar{s}, \alpha_M \bar{s}) \in Z \setminus \{0\}.$$

Jednocześnie $\alpha(\bar{s}, \bar{\eta}) = \alpha_M$ oraz $\frac{\bar{\eta}}{\|\bar{\eta}\|} = \bar{s} \in S_+^n(1)$, więc $\bar{s} \in S$.

(Wypukłość) Weźmy dowolne wektory $s^1, s^2 \in S$ oraz liczby $\alpha, \beta \geq 0$, $\alpha + \beta = 1$. Pokażemy, że $s = \alpha s^1 + \beta s^2 \in S$. Ponieważ $s^1, s^2 \in S$, więc

$$\exists (\bar{x}^i, \bar{y}^i) \in Z_{opt} \left(s^i = \frac{\bar{y}^i}{\|\bar{y}^i\|} \right), \; i = 1, 2.$$

Oczywiście, $\alpha(\bar{x}^1, \bar{y}^1) = \alpha(\bar{x}^2, \bar{y}^2) = \alpha_M$. Z **(G1)** oraz dodatniej jednorodności stopnia 0 funkcji α wynika, że $(\tilde{x}^i, s^i) \in Z_{opt}$, gdzie $\tilde{x}^i = \lambda_i \bar{x}^i$, $\lambda_i = \frac{1}{\|\bar{y}^i\|} > 0$, $i = 1, 2$. Stożek Z_{opt} jest zbiorem wypukłym, więc $(x', y') = \alpha(\tilde{x}^1, s^1) + \beta(\tilde{x}^2, s^2) \in Z_{opt}$ oraz

$$s = \alpha s^1 + \beta s^2 = \frac{\alpha s^1 + \beta s^2}{\|\alpha s^1 + \beta s^2\|} = \frac{y'}{\|y'\|}.$$

Zatem $s \in S$.

Zauważmy, że przy dowodzie zwartości i wypukłości zbioru *S* nie korzystamy z warunku **(G7)**.
(ii) Teza jest bezpośrednią konsekwencją warunku **(G7)**. ■

3. WIELOPASMOWA MAGISTRALA PRODUKCYJNA I RÓWNOWAGA VON NEUMANNA

Weźmy dowolny wektor (optymalnej) struktury produkcji $s \in S$. W literaturze półprostą

$$N_s = \{\lambda s | \lambda > 0\} \tag{7}$$

nazywa się promieniem von Neumanna (lub magistralą produkcyjną), zob. Takayama (1985, rozdz. 7), Nikaido (1968, rozdz. IV). Zbiór

$$\mathbb{N} = \bigcup_{s \in S} N_s = \{\lambda s |\ \lambda > 0, s \in S\} \tag{8}$$

nazywamy **wielopasmową magistralą produkcyjną** w gospodarce Gale'a. Magistralę wielopasmową $\mathbb{N}$ tworzy wiązka promieni (pojedynczych magistral) N_S, $s \in S$. Zbiór S składa się z wszystkich unormowanych (do 1) wektorów charakteryzujących strukturę produkcji na wielopasmowej magistrali.

□ **Twierdzenie 3.** Przestrzeń metryczna ($\mathbb{N}$, ρ) z metryką

$$\rho(N_{s^1}, N_{s^2}) = \left\| \frac{y^1}{\|y^1\|} - \frac{y^2}{\|y^2\|} \right\|, \ y^1 \in N_{s^1}, y^2 \in N_{s^2},$$

oraz operacjami dodawania:

$$N_{s^1} + N_{s^2} = \{y^1 + y^2 |\ y^1 \in N_{s^1},\ y^2 \in N_{s^2}\}$$

i mnożenia przez skalar $\sigma \in R_+^1$:

$$\sigma N_s = \{\sigma y | y \in N_s\} = \{\sigma\lambda s | \lambda > 0\}$$

jest zwarta i wypukła.

Dowód. (Zwartość) Niech $N_{s^i} \in \mathbb{N}$, i = 1, 2, …, ∞, tzn. $s^i \in S$, i = 1, 2, …, ∞. Zbiór S jest zwarty, a odwzorowanie $f: S \to \mathbb{N}$ postaci

$$f(s) = N_s = \{\lambda s | \lambda > 0\}$$

ciągłe, więc przestrzeń metryczna ($\mathbb{N}$, ρ) jest zwarta (jako ciągły obraz zbioru zwartego).

(Wypukłość) Weźmy dowolne promienie $N_{s^1}, N_{s^2} \in \mathbb{N}$ oraz liczby $\alpha, \beta \geq 0$, $\alpha + \beta = 1$. Niech $N = \alpha N_{s^1} + \beta N_{s^2}$. Pokażemy, że $N \in \mathbb{N}$. Jeżeli $\alpha = 0$, to $N = N_{s^2} \in \mathbb{N}$. Podobnie, jeżeli $\beta = 0$, to $N = N_{s^1} \in \mathbb{N}$. Załóżmy, że α, $\beta > 0$. Wówczas:

$$N = \alpha N_{s^1} + \beta N_{s^2} = \alpha\{\lambda s^1 | \lambda > 0\} + \beta\{\lambda s^2 | \lambda > 0\} = \{\lambda(\alpha s^1 + \beta s^2) | \lambda > 0\} \in \mathbb{N},$$

gdyż $s = \alpha s^1 + \beta s^2 \in S$. ■

Oznaczmy przez $p = (p_1, \ldots, p_n) \geq 0$ wektor cen towarów w gospodarce Gale'a. Niech $(x, y) \in Z$. Iloczyn $\langle p, x\rangle = \sum_{i=1}^{n} p_i x_i$ przedstawia wartość nakładów, a $\langle p, y\rangle = \sum_{i=1}^{n} p_i y_i$ wartość produkcji w procesie (x, y) (przy cenach p). Liczbę

$$\beta(x, y, p) = \frac{\langle p, y\rangle}{\langle p, x\rangle}$$

(tam gdzie $\langle p, x\rangle \neq 0$) nazywamy wskaźnikiem ekonomicznej efektywności procesu (x, y) (przy cenach p).

□ **Twierdzenie 4.** Przy założeniach **(G1)**–**(G7)** istnieje taki wektor cen $\bar{p} \geq 0$, że

$$\forall (x, y) \in Z(\langle \bar{p}, y\rangle - \alpha_M \langle \bar{p}, x\rangle \leq 0), \tag{9}$$

oraz

$$\forall (\bar{x}, \bar{y}) \in Z_{opt}(\langle \bar{p}, \bar{y}\rangle = \alpha_M \langle \bar{p}, \bar{x}\rangle > 0), \tag{10}$$

lub inaczej

$$\beta(x, y, \bar{p}) \leq \alpha_M \tag{9'}$$

(wszędzie gdzie funkcja $\beta(\cdot\,\cdot, \bar{p})$ jest określona) oraz

$$\beta(\bar{x}, \bar{y}, \bar{p}) = \alpha_M \tag{10'}$$

dla wszystkich procesów $(\bar{x}, \bar{y}) \in Z_{opt}$.

Dowód[6]**.** Zbiór

$$C = \{c \in R^n \mid c = \alpha_M x - y, (x, y) \in Z\}$$

jest stożkiem wypukłym w R^n (jako liniowy obraz stożka Z) nie zawierającym wektorów ujemnych. Istotnie, gdyby istniał taki proces $(x', y') \in Z$, że $c' = \alpha_M x' - y' < 0$, to znalazłaby się również taka liczba $\varepsilon' > 0$, że prawdziwa byłaby nierówność $(\alpha_M + \varepsilon')x' \leqq y'$. Wówczas musiałby zachodzić także warunek:

$$\alpha_M = \max_{(x,y)\in Z\setminus\{0\}} \alpha(x, y) \geq \alpha(x', y') \geq \alpha_M + \varepsilon',$$

co jest oczywiście niemożliwe. Zbiór

[6] Dowód nawiązuje do tw. 5.4 z pracy Panek (2003); zob. także Panek (2015a, tw. 1).

$$D = C + R_+^n = \{c + x | c \in C, x \in R_+^n\}$$

też jest stożkiem wypukłym z wierzchołkiem w 0 (jako suma pary stożków) nadal nie zawierającym wektorów ujemnych (nie ma ich bowiem w żadnym ze zbiorów C, R_+^n) oraz $C \subset D$. Natomiast do D należą wektory jednostkowe e^i = (0, …, 1, …, 0)), i = 1, 2, …, n. Z twierdzenia o hiperpłaszczyźnie oddzielającej wnioskujemy, że istnieje taki wektor $\bar{p} \neq 0$, że

$$\forall d \in D(\langle \bar{p}, d \rangle \geq 0).$$

Ponieważ $e^i \in D$, i = 1, 2, …, n, więc $\bar{p}_i \geq 0$, i = 1, 2, …, n. W szczególności

$$\forall c \in C(\langle \bar{p}, c \rangle \geq 0),$$

czyli

$$\forall (x, y) \in Z(\langle \bar{p}, \alpha_M x - y \rangle \geq 0),$$

tzn. zachodzi warunek (9) (równoważnie (9')).

Jeżeli $(\bar{x}, \bar{y}) \in Z_{opt}$, wtedy $\alpha_M \bar{x} \leqq \bar{y}$, więc

$$\langle \bar{p}, \bar{y} \rangle \geq \alpha_M \langle \bar{p}, \bar{x} \rangle.$$

Natomiast z (9), zważywszy na **(G7)** dostajemy:

$$0 < \langle \bar{p}, \bar{y} \rangle \leq \alpha_M \langle \bar{p}, \bar{x} \rangle.$$

Tym samym spełniony jest warunek (10) (równoważnie (10')).

Wektor $\bar{p}$ nazywamy wektorem cen von Neumanna. O każdej trójce $\{\alpha_{M,}(\bar{x}, \bar{y}), \bar{p}\}$ spełniającej warunki (9)–(10) mówimy, że tworzy (optymalny) stan równowagi von Neumanna[7].

4. DOPUSZCZALNE I STACJONARNE PROCESY WZROSTU

Zakładamy czas skokowy, t = 1, 2, …. Przez T = {0, 1, …, t_1}, $t_1 < +\infty$, oznaczamy horyzont funkcjonowania gospodarki, $x(t) = \big(x_1(t), \ldots, x_n(t)\big)$ jest wektorem

[7] W równowadze neumannowskiej ceny $\bar{p} \geq 0$ oraz proces produkcji $(\bar{x}, \bar{y}) \in Z \setminus \{0\}$ są określone z dokładnością do struktury (mnożenia przez stałą dodatnią). Dochodzi w niej do zrównania ekonomicznej efektywności produkcji z efektywnością technologiczną na maksymalnym możliwym do osiągnięcia przez gospodarkę poziomie.

nakładów (zużycia) towarów w okresie t, $y(t) = \big(y_1(t), \dots, y_n(t)\big)$ wektorem produkcji wytworzonej w tym okresie z nakładów $x(t)$. O parze wektorów $\big(x(t), y(t)\big) \in Z$ mówimy, że opisuje technologicznie dopuszczalny proces produkcji w okresie t. Gospodarka jest zamknięta w tym znaczeniu, że nakłady w okresie następnym, $t + 1$, pochodzą w niej wyłącznie z produkcji wytworzonej w okresie poprzednim t:

$$x(t+1) \leqq y(t),\ t = 0,\ 1,\ \dots,\ t_1 - 1,$$

co w myśl **(G4)** prowadzi do inkluzji:

$$\big(y(t), y(t+1)\big) \in Z,\ t = 0,\ 1,\ \dots,\ t_1 - 1. \tag{11}$$

Przez y^0 oznaczamy początkowy wektor produkcji:

$$y(0) = y^0 \geq 0. \tag{12}$$

O ciągu wektorów $\{y(t)\}_{t=0}^{t_1}$ spełniającym warunki (11)–(12) mówimy, że opisuje $(y^0,\ t_1)$ – dopuszczalny proces wzrostu w gospodarce Gale'a. Łatwo zauważyć, że przy przyjętych założeniach, dla dowolnego początkowego wektora produkcji $y^0 \geq 0$, w każdym (dowolnej długości) horyzoncie T istnieją $(y^0,\ t_1)$ – dopuszczalne procesy wzrostu; zob. np. Panek (2003, lemat 5.1). Proces dopuszczalny szczególnej postaci

$$y(t) = \gamma^t y^0,\ t = 0,\ 1,\ \dots,\ t_1 \tag{13}$$

$(\gamma > 0)$ nazywamy stacjonarnym procesem wzrostu z tempem γ. Nie zawsze (nie dla każdego wektora $y^0 \geq 0$) istnieje stacjonarny proces wzrostu. Proces taki (z tempem $\gamma > 0$) istnieje, gdy zachodzi inkluzja

$$(y^0, \gamma y^0) \in Z \setminus \{0\}.$$

Przy założeniach **(G1)**–**(G7)** w gospodarce Gale'a istnieją stacjonarne procesy wzrostu z dowolnym tempem $\gamma \in (0, \alpha_M]$:

$$\forall \gamma \in (0, \alpha_M] \exists \tilde{y} \geq 0 \big((\tilde{y}, \gamma\tilde{y}) \in Z \setminus \{0\}\big).$$

Zauważmy, że

$$\max\nolimits_{(y,\gamma y)\in Z\setminus\{0\}} \gamma = \max_{(x,y)\in Z\setminus\{0\}} \alpha(x,y) = \alpha(\bar{x}, \bar{y}).$$

Stąd, wobec **(G4)**, **(G5)** dostajemy:

$$(\bar{y}, \alpha_M \bar{y}) \in Z_{opt} \subset Z\setminus\{0\},$$

co prowadzi do stacjonarnego procesu $\{\bar{y}(t)\}_{t=0}^{t_1}$ postaci (13) z tempem $\gamma = \alpha_M$ i początkowym wektorem produkcji $y^0 = \bar{y}$, $\frac{\bar{y}}{\|\bar{y}\|} = s \in S$. Nazywamy go optymalnym stacjonarnym procesem wzrostu w gospodarce Gale'a. Optymalny stacjonarny proces postaci $\bar{y}(t) = \alpha_M^t \bar{y}$, $t = 0, 1, \ldots, t_1$ (z początkowym wektorem produkcji $\bar{y}$), biegnie po magistrali (promieniu) N_S:

$$\forall t \in T(\bar{y}(t) \in N_s),$$

gdzie $s = \frac{\bar{y}}{\|\bar{y}\|}$. Jego dowolna dodatnia λ – wielokrotność $\lambda\{\bar{y}(t)\}_{t=0}^{t_1}$ też jest optymalnym stacjonarnym procesem (z początkowym wektorem produkcji $\lambda\bar{y}$). Suma dwóch optymalnych stacjonarnych procesów $\bar{y}^1(t) = \alpha_M^t \bar{y}^1$, $\bar{y}^2(t) = \alpha_M^t \bar{y}^2$ jest optymalnym stacjonarnym procesem wzrostu (z początkowym wektorem produkcji $\bar{y} = \bar{y}^1 + \bar{y}^2$). Reasumując, optymalne stacjonarne procesy wzrostu w gospodarce Gale'a z wielopasmową magistralą tworzą stożek wypukły nie zawierający 0.

Jak pisaliśmy na wstępie, w pracach z teorii magistral przyjmuje się zazwyczaj, że optymalny stacjonarny proces jest w gospodarce Gale'a określony jednoznacznie z dokładnością do struktury (mnożenia przez stałą dodatnią), co skutkuje istnieniem dokładnie jednej magistrali N_S (jednego promienia von Neumanna), a nie wiązki magistral $\mathbb{N}$. Jednoznaczność promienia von Neumanna zapewnia następujący warunek[8]: jeżeli $\alpha_M \bar{x} = \bar{y} > 0$ oraz $s = \frac{\bar{y}}{\|\bar{y}\|}$, to

$$\forall (x, y) \in Z \setminus \{0\} \left(\frac{x}{\|x\|} \neq s \Rightarrow \beta(x, y, \bar{p}) = \frac{\langle \bar{p}, y \rangle}{\langle \bar{p}, x \rangle} < \alpha_M \right)$$

(jeżeli struktura nakładów w jakimkolwiek dopuszczalnym procesie produkcji odbiega od *jedynego* wektora *s* struktury produkcji na magistrali, to efektywność ekonomiczna takiego procesu jest niższa od optymalnej). W konsekwencji mamy jednoelementowy zbiór *S* oraz jeden, określony z dokładnością o struktury, optymalny proces produkcji $(\bar{x}, \bar{y}) \in Z_{opt}$.

Założenie o istnieniu w gospodarce tylko jednej magistrali nie ma uzasadnienia. Ponieważ nie można wykluczyć istnienia wielu optymalnych stacjonarnych ścieżek/procesów wzrostu, dlatego dalej zakładamy, że w gospodarce Gale'a może istnieć cała wiązka (pasmo) magistral $\mathbb{N}$. W myśl twierdzeń 2, 3 zbiór *S* wektorów struktury produkcji we wszystkich optymalnych procesach produkcji $(\bar{x}, \bar{y}) \in Z_{opt}$ oraz wielopasmowa magistrala $\mathbb{N}$, złożona z wiązki promieni von Neumanna $N_s, s \in S$, są zwarte i wypukłe. Po wielopasmowej magistrali biegną wszystkie optymalne stacjonarne procesy wzrostu postaci $\bar{y}(t) = \alpha_M^t \bar{y} > 0$, $\frac{\bar{y}}{\|\bar{y}\|} = s \in S$, $\left(\bar{y}(t), \bar{y}(t+1)\right) \in Z$, więc

[8] Zob. np. Radner (1961), Nikaido (1968, rozdz. IV, par. 13).

$$\beta(\bar{y}(t), \bar{y}(t+1), \bar{p}) = \frac{\langle \bar{p}, \bar{y}(t+1)\rangle}{\langle \bar{p}, \bar{y}(t)\rangle} = \frac{\langle \bar{p}, \alpha_M^{t+1}\bar{y}\rangle}{\langle \bar{p}, \alpha_M^{t}\bar{y}\rangle} = \alpha_M,\ t = 0,\ 1,\ \ldots,\ t_1 - 1,$$

co pokazuje, że procesy takie wyróżniają się także najwyższą efektywnością ekonomiczną. Wprowadźmy następującą miarę odległości (kątowej) wektora $x = (x_1, \ldots, x_n) \geq 0$ od wielopasmowej magistrali (zbioru) $\mathbb{N}$:

$$d(x, \mathbb{N}) = \inf_{x' \in \mathbb{N}} \left\| \frac{x}{\|x\|} - \frac{x'}{\|x'\|} \right\|.$$

Zakładamy, że

(G8) $\forall (x, y) \in Z \setminus \{0\} (x \notin \mathbb{N} \Rightarrow \beta(x, y, \bar{p}) < \alpha_M)$

(efektywność ekonomiczna procesu leżącego poza wielopasmową magistralą jest niższa od optymalnej) lub równoważnie:

(G8') $\forall (x, y) \in Z \setminus \{0\} (x \notin \mathbb{N} \Rightarrow \langle \bar{p}, y\rangle < \alpha_M \langle \bar{p}, x\rangle)$.

Warunek **(G8)** nie wyklucza istnienia wielu magistral, tym samym występowania w równowadze von Neumanna wielu optymalnych procesów produkcji o różnej strukturze. Wszystkie wyróżnia też najwyższa efektywność ekonomiczna.

□ **Twierdzenie 5.** Jeżeli zachodzą warunki **(G1)**–**(G8)**, to

$$\forall \varepsilon > 0\ \exists \delta_\varepsilon \in (0, \alpha_M) \forall (x, y) \in Z \setminus \{0\}\ (d(x, \mathbb{N}) \geq \varepsilon \Rightarrow \beta(x, y, \bar{p}) \leq \alpha_M - \delta_\varepsilon).$$

Dowód[9]**.** Zauważmy, że jeżeli proces $(x, y) \in Z \setminus \{0\}$ spełnia warunki twierdzenia, to spełnia je także każdy proces $\lambda(x, y)$ z liczbą $\lambda > 0$. Przy dowodzie wystarczy ograniczyć się do dopuszczalnych procesów produkcji (x, y) ze zbioru

$$V(\varepsilon) = \{(x, y) \in Z | \|x\| = 1 \wedge d(x, \mathbb{N}) \geq \varepsilon\}.$$

Z definicji mamy $d(x, \mathbb{N}) = \inf_{x' \in \mathbb{N}} f(x, x')$, gdzie $f(x, x') = \left\| \frac{x}{\|x\|} - \frac{x'}{\|x'\|} \right\|$, lub inaczej:

$$d(x, \mathbb{N}) = \inf_{s \in S} f(x, s),\ s = \frac{x'}{\|x'\|}.$$

[9] Dowód jest (dostosowaną do specyfiki obecnego modelu) wersją lematu Radnera (1961). Odległość (kątową) wektora x od pojedynczej magistrali $N_s = \{\lambda s | \lambda > 0\}$ postaci $d(x, N) = \left\| \frac{x}{\|x\|} - s \right\|$ w oryginalnym dowodzie Radnera zastępuje obecnie odległość $d(x, \mathbb{N})$ wektora x od zbioru (wiązki) magistral $\mathbb{N}$. Technika dowodzenia pozostaje bez zmian.

Ponieważ $f \in C^0(R_+^n \setminus \{0\} \times R_+^n \setminus \{0\})$ oraz zbiór S jest zwarty, więc

$$\forall x \geq 0 \; \exists \bar{s} \in S \left(f(x, \bar{s})\right) = \inf_{s \in S} f(x, s). \tag{14}$$

Ustalmy dowolną liczbę $\varepsilon > 0$. Pokażemy, że zbiór $V(\varepsilon)$ jest zwarty (ograniczony i domknięty w R^{2n}).

(Ograniczoność) Załóżmy, że $\exists \left\{x^i, y^i\right\}_{i=1}^{\infty}, \left(x^i, y^i\right) \in V(\varepsilon), \; \left\|\left(x^i, y^i\right)\right\| \underset{i}{\to} +\infty$. Wobec tego, że $\forall i \left(\left\|x^i\right\| = 1\right)$ wnioskujemy, iż $\left\|y^i\right\| \underset{i}{\to} +\infty$. Wtedy, zgodnie z **(G1)**, istnieje ciąg dopuszczalnych procesów

$$\left(\xi^i, \eta^i\right) = \left(\frac{x^i}{\|y^i\|}, \frac{y^i}{\|y^i\|}\right) \in Z, \; i = 1, 2, \ldots,$$

$\xi^i \underset{i}{\to} 0, \; \left\|\eta^i\right\| = 1$, zatem istnieje podciąg $\left\{\xi^{i_j}, \eta^{i_j}\right\}_{j=1}^{\infty}$ zbieżny do granicy $(0, \bar{\eta})$ oraz $\|\bar{\eta}\| = 1$. Przestrzeń produkcyjna Z jest zbiorem domkniętym w R^{2n}, więc $(0, \bar{\eta}) \in Z$, co przeczy jednak **(G3)**. Tym samym $V(\varepsilon)$ jest zbiorem ograniczonym.

(Domkniętość) Niech $\left\{x^i, y^i\right\}_{i=1}^{\infty}$ będzie ciągiem procesów ze zbioru $V(\varepsilon)$ zbieżnym do granicy $(\bar{x}, \bar{y})$. Wówczas oczywiście $(\bar{x}, \bar{y}) \in Z$ oraz $\|\bar{x}\| = 1$. Pokażemy, że $d(\bar{x}, \mathbb{N}) \geq \varepsilon$. Wobec ciągłości funkcji $f(\cdot, \cdot)$ i zwartości zbioru S, zgodnie z (14):

$$\forall i \; \exists s^i \in S \left(f\left(x^i, s^i\right) = \inf_{s \in S} f\left(x^i, s\right) = d\left(x^i, \mathbb{N}\right) \geq \varepsilon\right)$$

oraz

$$\exists \left\{s^{i_j}\right\}_{j=1}^{\infty} \left(s^{i_j} \underset{j}{\to} \bar{s} \in S \wedge f\left(x^{i_j}, s^{i_j}\right) = \inf_{s \in S} f\left(x^{i_j}, s\right) = d\left(x^{i_j}, \mathbb{N}\right) \geq \varepsilon\right).$$

Wówczas $f(\bar{x}, \bar{s}) = d(\bar{x}, \mathbb{N}) \geq \varepsilon$. Zbiór $V(\varepsilon)$ jest zatem domknięty w R^{2n}. Jest także ograniczony, więc jest zwarty.

W myśl **(G8')**

$$\forall (x, y) \in V(\varepsilon) (0 \leq \langle \bar{p}, y \rangle < \alpha_M \langle \bar{p}, x \rangle),$$

czyli

$$\forall (x, y) \in V(\varepsilon) \left(0 \leq \beta(x, y, \bar{p}) = \frac{\langle \bar{p}, y \rangle}{\langle \bar{p}, x \rangle} < \alpha_M\right).$$

Funkcja $\beta(\cdot,\cdot,\bar{p})$ jest ciągła na $V(\varepsilon)$ jako iloraz dwóch funkcji liniowych z niezerową funkcją w mianowniku (wszędzie na $V(\varepsilon)$ mamy bowiem $\langle \bar{p}, x\rangle > 0$), więc w myśl twierdzenia Weierstrassa istnieje rozwiązanie zadania

$$\max_{(x,y)\in V(\varepsilon)} \beta(x,y,\bar{p}) = \beta_\varepsilon < \alpha_M.$$

Wówczas

$$\exists \delta_\varepsilon > 0 \forall (x,y) \in V(\varepsilon)(\beta(x,y,\bar{p}) \leq \alpha_M - \delta_\varepsilon),$$

(wystarczy przyjąć $\delta_\varepsilon = \alpha_M - \beta_\varepsilon > 0$) lub równoważnie:

$$\forall (x,y) \in V(\varepsilon)(\langle \bar{p}, y\rangle \leq (\alpha_M - \delta_\varepsilon)\langle \bar{p}, x\rangle),$$

c.n.d. ■

5. OPTYMALNE PROCESY WZROSTU. „SŁABE" TWIERDZENIE O WIELOPASMOWEJ MAGISTRALI

Interesuje nas następujące zadanie maksymalizacji wartości produkcji mierzonej w cenach von Neumanna w końcowym okresie t_1 horyzontu T[10]:

$$\max \langle \bar{p}, y(t_1)\rangle$$

$$\text{p.w. } (11)-(12).$$

Przy założeniach **(G1)**–**(G8)** zadanie to ma rozwiązanie $\{y^*(t)\}_{t=0}^{t_1}$, które nazywamy $(y^0, t_1, \bar{p})$ – optymalnym procesem wzrostu (trajektorią produkcji) w gospodarce Gale'a, zob. np. Panek (2003, tw. 5.7). Kolejne twierdzenie (twierdzenie 6) mówi o specyficznej stabilności optymalnych procesów wzrostu w gospodarce Gale'a w otoczeniu wielopasmowej magistrali $\mathbb{N}$. Decydujemy się na przedstawienie jego dowodu, mimo że jest on wzorowany na dowodach podobnych twierdzeń w gospodarce Gale'a z pojedynczą magistralą, przedstawionych przez autora, m.in. w pracach (2003, tw. 5.8; 2013, tw. 4). W szczegółach różni się na tyle istotnie, iż uznaliśmy jego przytoczenie za zasadne.

[10] Tzw. zagadnienie wzrostu docelowego.

□ **Twierdzenie 6** („Słabe" twierdzenie o magistrali). Jeżeli zachodzą warunki **(G1)**–**(G8)** oraz istnieje taki $(y^0, \check{t})$ – dopuszczalny proces $\{\check{y}(t)\}_{t=0}^{\check{t}}$, $\check{t} < t_1$, że $\check{y}(\check{t}) \in \mathbb{N}$, to dla dowolnej liczby $\varepsilon > 0$ istnieje taka liczba naturalna k_ε, że liczba okresów czasu, w których $(y^0, t_1, \bar{p})$ – optymalny proces wzrostu $\{y^*(t)\}_{t=0}^{t_1}$ spełnia warunek:

$$d(y^*(t), \mathbb{N}) \geq \varepsilon \tag{15}$$

nie przekracza k_ε. Liczba k_ε nie zależy od długości horyzontu T.

Dowód. Z definicji $(y^0, t_1, \bar{p})$ – optymalnego procesu wzrostu $\{y^*(t)\}_{t=0}^{t_1}$, zgodnie z (9), mamy:

$$\langle \bar{p}, y^*(t+1) \rangle \leq \alpha_M \langle \bar{p}, y^*(t) \rangle, \; t = 0, 1, \ldots, t_1 - 1. \tag{16}$$

Załóżmy, że w okresach $\tau_1, \ldots, \tau_k < t_1$ zachodzi warunek (15). Wówczas, w myśl twierdzenia 5 istnieje taka liczba $\delta_\varepsilon \in (0, \alpha_M)$, że

$$\langle \bar{p}, y^*(t+1) \rangle \leq (\alpha_M - \delta_\varepsilon) \langle \bar{p}, y^*(t) \rangle, \; t = \tau_1, \ldots, \tau_k. \tag{17}$$

Łącząc (16), (17) dostajemy:

$$\langle \bar{p}, y^*(t_1) \rangle \leq \alpha_M^{t_1 - k} (\alpha_M - \delta_\varepsilon)^k \langle \bar{p}, y^0 \rangle. \tag{18}$$

Z założenia istnieje $(y^0, \check{t})$ – dopuszczalny proces $\{\check{y}(t)\}_{t=0}^{\check{t}}$ prowadzący w okresie $\check{t} < t_1$ do wielopasmowej magistrali $\mathbb{N}$, zatem istnieje liczba $\sigma > 0$ i taki wektor $s \in S$, że $\sigma s = \check{y}(\check{t})$ oraz (y^0, t_1) – dopuszczalny jest proces $\{\tilde{y}(t)\}_{t=0}^{t_1}$ postaci

$$\tilde{y}(t) = \begin{cases} \check{y}(t), t = 0, 1, \ldots, \check{t}, \\ \sigma s \alpha_M^{t - \check{t}}, \; t = \check{t} + 1, \ldots, t_1 \end{cases} \tag{19}$$

utworzony ze „sklejenia" procesu $\{\check{y}(t)\}_{t=0}^{\check{t}}$ i stacjonarnego procesu postaci (13) z tempem α_M (dla $t > \check{t}$). Wówczas

$$\langle \bar{p}, y^*(t_1) \rangle \geq \langle \bar{p}, \tilde{y}(t_1) \rangle = \sigma \alpha_M^{t_1 - \check{t}} \langle \bar{p}, s \rangle > 0. \tag{20}$$

Z (18), (20) wynika nierówność:

$$\sigma \alpha_M^{t_1 - \check{t}} \langle \bar{p}, s \rangle \leq \alpha_M^{t_1 - k} (\alpha_M - \delta_\varepsilon)^k \langle \bar{p}, y^0 \rangle,$$

a stąd, po przekształceniach otrzymujemy:

$$k \leq \frac{\ln A}{\ln \alpha_M - \ln(\alpha_M - \delta_\varepsilon)} = B, \tag{21}$$

gdzie $0 < A = \max_{s \in S} \frac{\alpha_M^{\check{t}} \langle \bar{p}, y^0 \rangle}{\sigma \langle \bar{p}, s \rangle} < +\infty$. W charakterze liczby k_ε wystarczy wziąć najmniejszą liczbę naturalną większą od $\max\{0, B+1\}$. ■

Twierdzenie orzeka, że jeżeli choćby jeden dopuszczalny proces wzrostu prowadzi ze stanu początkowego y^0 do wielopasmowej magistrali $\mathbb{N}$, to w długich okresach czasu każdy optymalny proces wzrostu „prawie zawsze" przebiega w jej dowolnie bliskim otoczeniu (kątowym).

6. UOGÓLNIENIA

Twierdzenie 4 o istnieniu optymalnego stanu równowagi von Neumanna $\{\alpha_M, (\bar{x}, \bar{y}), \bar{p}\}$ pozostaje prawdziwe po zastąpieniu warunku silnej regularności **(G7)** następującym warunkiem słabej regularności gospodarki Gale'a:

(G7') $\forall s \in S\ \exists t_s < +\infty \wedge\ \exists (s, t_s)$ – dopuszczalny proces $\{y(t)\}_{t=0}^{t_s}$ spełniający warunek:

$$\big(y(t), y(t+1)\big) \in Z,\ t = 0,\ 1,\ \ldots,\ t_s - 1,$$

$$y(0) = s, y(t_s) > 0.$$

Mówi o tym twierdzenie 4'. Każdy (unormowany do 1) wektor $s \in S$ spełniający warunek **(G7)** jest dodatni (zob. twierdzenie 2(ii)), natomiast obecnie zbiór S może zawierać zarówno wektory dodatnie jak i półdodatnie (nieujemne i niezerowe)[11]. Warunek **(G7')** jest równoważny z warunkiem głoszącym, że z każdego miejsca (punktu) wielopasmowej magistrali $\mathbb{N}$ pewien proces wzrostu w skończonym czasie t_s prowadzi do dodatniego wektora produkcji $y(t_s)$. Oczywiście, warunek ten jest słabszy od **(G7)**, tzn. jeżeli zachodzi **(G7)**, to zachodzi także **(G7')** (dla $t_s = 0$).

□ **Twierdzenie 4'.** Przy założeniach **(G1)**–**(G6)**, **(G7')** zachodzą warunki (9)–(10) (równoważnie (9')–(10')).

Dowód warunku (9) nie zmienia się (przebiega analogicznie jak w twierdzeniu 4). Przechodząc do dowodu warunku (10) zauważmy, że wprawdzie przy założeniach

[11] Odnosi się to rzecz jasna także do magistral, na których mogą obecnie znajdować się nie tylko dodatnie, ale również półdodatnie wektory produkcji. Zwracamy także uwagę, że po zastąpieniu warunku **(G7)** warunkiem **(G7')** zbiór S pozostaje zwarty; zob. dowód twierdzenia 2(i).

(G1)–(G6) każdy optymalny proces produkcji $(\bar{x}, \bar{y}) \in Z_{opt}$ spełnia nierówność $0 \leq \alpha_M \bar{x} \leqq \bar{y}$ i wobec tego

$$\forall (\bar{x}, \bar{y}) \in Z_{opt}(\langle \bar{p}, \bar{y} \rangle \geq \alpha_M \langle \bar{p}, \bar{x} \rangle \geq 0),$$

ale wektor produkcji $\bar{y}$ nie musi być dodatni. Z drugiej strony, w myśl (9)

$$\forall (\bar{x}, \bar{y}) \in Z_{opt} \subset Z(\langle \bar{p}, \bar{y} \rangle \leq \alpha_M \langle \bar{p}, \bar{x} \rangle),$$

zatem każdy proces produkcji $(\bar{x}, \bar{y}) \in Z_{opt}$ spełnia warunek:

$$\langle \bar{p}, \bar{y} \rangle = \alpha_M \langle \bar{p}, \bar{x} \rangle. \tag{22}$$

Wobec (9) i **(G7')** $\forall s \in S$ istnieje taki (s, t_s) – dopuszczalny proces $\{y(t)\}_{t=0}^{t_s}$, że

$$0 < \langle \bar{p}, y(t_s) \rangle \leq \alpha_M \langle \bar{p}, y(t_s - 1) \rangle \leq \ldots \leq \alpha_M^{t_s} \langle \bar{p}, y(0) \rangle = \alpha_M^{t_s} \langle \bar{p}, s \rangle, \tag{23}$$

gdzie $s = \frac{\bar{y}}{\|\bar{y}\|}$, czyli (zważywszy na definicję zbioru *S*; zob. (6)):

$$\forall (\bar{x}, \bar{y}) \in Z_{opt}(\langle \bar{p}, \bar{y} \rangle > 0).$$

Stąd i z (22) otrzymujemy warunek (10) (równoważnie (10')). ■

„Słabe" twierdzenie o magistrali (twierdzenie 6) zachowuje moc, gdy warunek **(G7)** zastąpimy słabszym warunkiem **(G7')**. Co więcej, zastąpienie w warunku początkowym (14) półdodatniego wektora produkcji y^0 wektorem dodatnim zapewnia istnienie $(y^0, \check{t})$ – dopuszczalnego procesu $\{\check{y}(t)\}_{t=0}^{\check{t}}$ prowadzącego do wielopasmowej magistrali $\mathbb{N}$ już w okresie $\check{t} = 1$. „Słabe" twierdzenie o magistrali przyjmuje wówczas postać następującą.

□ **Twierdzenie 6'.** Jeżeli gospodarka Gale'a z dodatnim początkowym wektorem produkcji y^0 spełnia warunki **(G1)–(G6)**, **(G7')**, **(G8)**, to dla dowolnej liczby $\varepsilon > 0$ istnieje taka liczba naturalna *k*, że liczba okresów czasu, w których $(y^0, t_1, \bar{p})$ – optymalny proces wzrostu $\{y^*(t)\}_{t=0}^{t_1}$ spełnia warunek (15) nie przekracza *k*. Liczba *k* nie zależy od długości horyzontu *T*.

Dowód jest dosłownym powtórzeniem dowodu twierdzenia 6, z tym że obecnie (y^0, t_1) – dopuszczalny proces wzrostu $\{\tilde{y}(t)\}_{t=0}^{t_1}$ przyjmuje postać następującą:

$$\tilde{y}(t) = \begin{cases} y^0, & t = 0, \\ \sigma s \alpha_M^t, & t = 1, \ldots, t_1, \end{cases}$$

gdzie $\sigma = \min_i \frac{y_i^0}{\bar{s}_i} > 0$. Zamiast warunku (20) mamy[12]:

$$\langle \bar{p}, y^*(t_1)\rangle \geq \langle \bar{p}, \tilde{y}(t_1)\rangle = \sigma \alpha_M^{t_1} \langle \bar{p}, s\rangle > 0. \qquad (20')$$

W ten sposób zastąpienie warunku **(G7)** słabszym warunkiem **(G7')** znowu prowadzi do oszacowania (21) i zamyka dowód. ■

Niech u: $R_+^n \to R_+^1$ będzie ciągłą, wklęsłą, rosnącą i dodatnio jednorodną stopnia 1 społeczną funkcją użyteczności produkcji, spełniającą dodatkowo następujące warunki:

(G9) (i) $\exists a > 0\ \forall\, y \geq 0\ (u(y) \leq a\,\langle \bar{p}, y\rangle)$, **(ii)** $\forall s \in S\ (u(s) > 0)$[13].

Warunek **(i)** stanowi, że funkcję użyteczności można aproksymować (z góry) formą liniową (z wektorem tworzącym $a\bar{p}$, gdzie a jest pewną liczbą dodatnią)[14]. Zgodnie z **(ii)** produkcję na magistrali charakteryzuje dodatnia użyteczność. Rozwiązanie $\{y^*(t)\}_{t=0}^{t_1}$ zadania

$$\max\ u(y(t_1))$$
$$\text{p.w. } (11)–(12)$$

nazywamy (y^o, t_1, u) – optymalnym procesem wzrostu w gospodarce Gale'a[15].

□ **Twierdzenie 7.** Jeżeli zachodzą warunki **(G1)**–**(G9)** oraz istnieje taki $(y^0, \check{t})$ – dopuszczalny proces $\{\check{y}(t)\}_{t=0}^{\check{t}}$, $\check{t} < t_1$, że $\check{y}(\check{t}) \in \mathbb{N}$, to dla dowolnej liczby $\varepsilon > 0$ istnieje taka liczba naturalna k, iż liczba okresów czasu, w których (y^0, t_1, u) – optymalny proces wzrostu $\{y^*(t)\}_{t=0}^{t_1}$ spełnia warunek (15) nie przekracza k. Liczba k zależy od ε oraz nie zależy od długości horyzontu T.

Dowód. Postępując analogicznie jak przy dowodzie twierdzenia 6 dochodzimy do warunku (18):

$$\langle \bar{p}, y^*(t_1)\rangle \leq \alpha_M^{t_1-k}(\alpha_M - \delta_\varepsilon)^k \langle \bar{p}, y^0\rangle.$$

Stąd, wobec **(G9)**, otrzymujemy ograniczenie górne (wartości funkcji użyteczności):

---

[12] Parametry σ oraz α_M są bowiem dodatnie oraz $\langle \bar{p}, s\rangle > 0$ (zgodnie z (23)).

[13] Jeżeli zachodzi warunek **(G7)**, to (ii) można zastąpić równoważnym warunkiem: $u(y) > 0$ choćby dla jednego wektora $y > 0$.

[14] Zob. np. Takayama (1985, rozdz. 7). Warunki takie spełniają np. dodatnio jednorodne stopnia 1 funkcje użyteczności typu Cobba-Douglasa i CES oraz funkcja Leontiefa-Koopmansa.

[15] Przy przyjętych założeniach procesy takie istnieją dla dowolnego wektora początkowego $y^0 \geq 0$ i dowolnej długości (skończonego) horyzontu T.

$$u\left(y^*(t_1)\right) \leq a\langle \bar{p}, y^*(t_1)\rangle \leq a\alpha_M^{t_1-k}(\alpha_M - \delta_\varepsilon)^k\langle \bar{p}, y^0\rangle, \tag{24}$$

gdzie *a* jest pewną liczbą dodatnią. Z drugiej strony, zważywszy na definicję (y^0, t_1, u) – optymalnego procesu wzrostu $\{y^*(t)\}_{t=0}^{t_1}$, własność **(i)**, **(ii)** funkcji użyteczności (warunek **(G9)**) oraz konstrukcję $(y^0, \check{t})$ – dopuszczalnego procesu $\{\check{y}(t)\}_{t=0}^{\check{t}}$ (zob. (19)), dochodzimy do ograniczenia dolnego:

$$u\left(y^*(t_1)\right) \geq u\left(\tilde{y}(t_1)\right) = u\left(\sigma s\alpha_M^{t_1-\check{t}}\right) = \sigma\alpha_M^{t_1-\check{t}}u(s) > 0. \tag{25}$$

Z (24), (25) otrzymujemy nierówność:

$$\sigma\alpha_M^{t_1-\check{t}}u(s) \leq a\alpha_M^{t_1-k}(\alpha_M - \delta_\varepsilon)^k\langle \bar{p}, y^0\rangle,$$

co daje oszacowanie (21) liczby *k*, z tym że obecnie $0 < A = \max_{s\in S}\frac{a\alpha_M^{\check{t}}\langle \bar{p}, y^0\rangle}{\sigma u(s)} < +\infty$. Podobnie jak w twierdzeniu 6, w charakterze liczby k_ε wystarczy wziąć najmniejszą liczbę naturalną większą od $\max\{0, B+1\}$, gdzie $B = \frac{\ln A}{\ln\alpha_M - \ln(\alpha_M - \delta_\varepsilon)}$. ■

Twierdzenie 7, podobnie jak twierdzenie 6, pozostaje prawdziwe, gdy warunek **(G7)** zastąpimy w nim warunkiem **(G7')**. Jeżeli ponadto przyjmiemy, że początkowy wektor produkcji y^0 jest dodatni, wtedy dostajemy wersję „słabego" twierdzenia o magistrali podobną do twierdzenia 6'.

7. UWAGI KOŃCOWE

Przedstawione twierdzenia (6, 7) nawiązują do podobnych twierdzeń mówiących o własnościach optymalnych procesów wzrostu w gospodarce Gale'a z pojedynczą magistralą produkcyjną (jednym promieniem von Neumanna)[16]. W artykule uchylamy ten warunek i dopuszczamy istnienie w gospodarce Gale'a zbioru magistral, złożonego ze zwartej wiązki promieni von Neumanna. Obydwa udowodnione twierdzenia o wielopasmowej magistrali należą do relatywnie najprostszej grupy tzw. „słabych" twierdzeń mówiących wprawdzie o zbieżności optymalnych procesów do magistrali, jednak bez wskazania okresów czasu, w których zbieżność taka zachodzi. Z tego punktu widzenia inspirujące i zachęcające do dalszych badań są tzw. „silne" oraz „bardzo silne" twierdzenia o wielopasmowej magistrali na gruncie modeli Neumanna-Gale'a.

[16] Zob. np. wspomniane wcześniej prace: Nikaido (1968, rozdz. IV, tw. 13.8), Panek (2003, rozdz. 5, tw. 5.8), Takayama (1985, rozdz. 7).

LITERATURA

McKenzie L. W., (1976), Turnpike Theory, *Econometrica*, 44, 841–866.
McKenzie L. W., (1998), Turnpikes, *American Economic Review*, 88 (2), 1–14.
McKenzie L. W., (2005), Optimal Economic Growth, Turnpike Theorems and Comparative Dynamics, w: Arrow K. J., Intriligator M. D., (red.), *Handbook of Mathematical Economics*, ed. 2, vol. III, chapter 26, 1281–1355.
Nikaido H., (1968), *Convex Structures and Economic Theory*, Acad. Press, New York.
Panek E., (2003), *Ekonomia matematyczna*, Wydawnictwo AE, Poznań.
Panek E., (2013), „Słaby" i „bardzo silny" efekt magistrali w niestacjonarnej gospodarce Gale'a z graniczną technologią, *Przegląd Statystyczny*, 60 (3), 291–303.
Panek E., (2014), Niestacjonarna gospodarka Gale'a z rosnącą efektywnością produkcji na magistrali, *Przegląd Statystyczny*, 61 (1), 5–14.
Panek E., (2015a), Zakrzywiona magistrala w niestacjonarnej gospodarce Gale'a. Część I, *Przegląd Statystyczny*, 62 (2), 149–163.
Panek E., (2015b), Zakrzywiona magistrala w niestacjonarnej gospodarce Gale'a. Część II, *Przegląd Statystyczny*, 62 (2), 349–360.
Panek E., (2016), „Silny" efekt magistrali w modelu niestacjonarnej gospodarki z graniczną technologią, *Przegląd Statystyczny*, 63 (2), 109–121.
Radner R., (1961), Paths of Economic Growth that are Optimal with Regard only to Final States: A Turnpike Theorem, *Review of Economic Studies*, 28 (2), 98–104.
Takayama A., (1985), *Mathematical Economics*, Cambridge University Press, Cambridge.

GOSPODARKA GALE'A Z WIELOMA MAGISTRALAMI. „SŁABY" EFEKT MAGISTRALI

Streszczenie

W bogatej literaturze z teorii magistral zazwyczaj zakłada się, że wzorcowa ścieżka – zwana magistralą – do której w długich okresach czasu zbieżne są wszystkie optymalne procesy wzrostu, jest określona jednoznacznie. W modelu Gale'a (w wersji stacjonarnej) jej obrazem geometrycznym jest półprosta w przestrzeni stanów gospodarki, zwana promieniem von Neumanna. W artykule uchylamy założenie jednoznaczności magistrali (promienia von Neumanna) i badamy zachowanie stacjonarnej gospodarki typu Gale'a ze zwartą wiązką magistral, którą umownie nazywamy magistralą wielopasmową. Prezentujemy dowód kilku wariantów „słabego" twierdzenia o wielopasmowej magistrali w stacjonarnej gospodarce Gale'a.

Słowa kluczowe: model wzrostu Gale'a, równowaga von Neumanna, warunek silnej (słabej) regularności, wielopasmowa magistrala produkcyjna, „słaby" efekt magistrali

GALE'S ECONOMY WITH MULTIPLE TURNPIKES. "WEAK" TURNPIKE EFFECT

Abstract

In the vast literature on turnpike theory it is generally assumed that the model path – called the turnpike – to which in a long time period all the optimal processes are convergent, is uniquely determined. Its geometric image in the Gale's model (in its stationary version) is a ray in the space of all states of

the economy. We call it von Neumann's ray. In this paper we evade the assumption of the uniqueness of this turnpike (von Neumann's ray) and study the behaviour of the stationary Gale's economy with the compact turnpikes' bundle. We call it multilane turnpike. We present proofs for several variants of the "weak" multilane turnpike theorem in the stationary Gales' economy.

Keywords: Gale economy, von Neumann equilibrium, strong (weak) regularity condition, multilane production turnpike, "weak" turnpike theorem

DOI: 10.5604/01.3001.0014.1215

PRZEGLĄD STATYSTYCZNY
R. LXIII – ZESZYT 4 – 2016

ŁUKASZ LENART[1], MATEUSZ PIPIEŃ[2]

KONCEPCJA WSTĘGOWEGO ZEGARA CYKLU KONIUNKTURALNEGO W UJĘCIU NIEPARAMETRYCZNYM[3]

1. WPROWADZENIE

Zegary cyklu koniunkturalnego są powszechnie stosowanym narzędziem w monitorowaniu i prezentacji aktualnej pozycji cyklicznej gospodarki. Narzędzie to jest z powodzeniem stosowane w wielu bankach centralnych oraz instytucjach, w których analiza aktualnej pozycji cyklu koniunkturalnego jest ważna w celu podjęcia optymalnych decyzji. Popularność zegara ma swe źródło w prostym w interpretacji sposobie wizualizacji zmian w cyklu koniunkturalnym, który odnosi się bezpośrednio do wykresu fazowego. Zegar jest stosowany między innymi przez EUROSTAT, sprawozdawczość statystyczną OECD, Statistics Netherlands, Statistisches Bundesamt Deutschland (patrz Abberger, Nierhaus, 2010) oraz wielu bankach centralnych.

Podejście, które jest wykorzystywane w konstrukcji zegara cyklu opiera się w głównej mierze na analizie dynamiki wyodrębnionych wahań, które utożsamia się z wahaniami aktywności gospodarczej. Wnioski jakie są wynikiem obserwacji dynamiki zegara są jednak bardzo często obarczone dużą niepewnością wynikającą z przyjętej metodologii wyodrębnienia komponentu cyklicznego. Mechanizm zegara cyklu koniunkturalnego oparty jest bowiem w głównej mierze na dynamice wyodrębnionych wahań utożsamianych z wahaniami aktywności gospodarczej. Wszystkie kwestie jakie dotyczą poprawnego wyodrębnienia wahań cyklicznych przenoszą się zatem automatycznie na poprawną dynamikę zegara cyklu. Jednym z kluczowych problemów podczas wyodrębniania wahań utożsamianych z wahaniami aktywności gospodarczej na podstawie wskaźników makroekonomicznych jest dobór parametrów. Jeśli metodą wyodrębniania jest filtr pasmowo-przepustowy wtedy różne wartości parametrów metody mogą wpływać znacząco na otrzymane wyniki. W przypadku

[1] Uniwersytet Ekonomiczny w Krakowie, Wydział Finansów, Katedra Matematyki, ul. Rakowicka 27, 31-510 Kraków, Polska.

[2] Uniwersytet Ekonomiczny w Krakowie, Wydział Zarządzania, Katedra Ekonometrii i Badań Operacyjnych, ul. Rakowicka 27, 31-510 Kraków, Polska, autor prowadzący korespondencję – e-mail: eepipien@cyf-kr.edu.pl.

[3] Badania finansowane przez Narodowe Centrum Nauki w ramach grantu OPUS, numer grantu DEC-2013/09/B/HS4/01945.

filtracji filtrem zaproponowanym przez Hodricka, Prescotta (1997) (w skrócie HP) kluczowy jest dobór parametru wygładzającego λ.

Celem pracy jest uwzględnienie w metodzie wyodrębniania cyklu filtrem HP przedziału dla parametru λ, zamiast pojedynczej arbitralnej wartości. W ten sposób interpretacji podlega nie arbitralnie wyodrębniony cykl dla jednego parametru wygładzającego a znacznie szersze spektrum wahań cyklicznych. W oparciu o tak otrzymane spektrum wahań cyklicznych skonstruowano nieparametryczny zegar koniunkturalny (wahań cyklicznych) uwzględniający przedział parametrów w metodzie HP (tzw. *wstęgowy zegar koniunkturalny*). Wahania skonstruowanego zegara zdekomponowano na wahania fazy cyklu oraz wahania amplitudy cyklu. W analizie fazy wahań cyklicznych również uwzględniono przedział parametrów wygładzających, tworząc tzw. *wstęgowy wykres fazy cyklu*.

W artkule przedstawiono też metodę ilustracji dwóch charakterystyk cykliczności, to jest fazy i amplitudy w sposób odseparowany. Wyodrębnienie fazy cyklu z pominięciem jego amplitudy może bowiem dostarczyć bardziej precyzyjnego określenia stanu koniunktury, bez wnikania w szczegóły dotyczące głębokości jej wahań. Wyniki empiryczne w ramach tego problemu wyraźnie wskazują na potrzebę określenia dodatkowo odrębnej fazy cyklu, zwaną przez autorów fazą neutralną.

Rozdział 2 przedstawia ideę konstrukcji zegara cyklu koniunkturalnego opartego na jednowymiarowych realnym wskaźniku makroekonomicznym. W rozdziale 3 zaproponowano konstrukcję nieparametrycznego zegara cyklu koniunkturalnego zgodnie z równaniem modelu wprowadzonym w pracy Lenart, Pipień (2013). W części empirycznej rozdziału wskazano na wrażliwość dynamiki zegara dla PKB Polski na różne wartości parametru wygładzającego metody HP. Kolejny rozdział przedstawia metodologię odseparowania fazy od wahań cyklicznych na podstawie dynamiki wstęgowego zegara cyklu. W części empirycznej tego rozdziału proces fazy wyodrębniono dla wcześniej analizowanego wskaźnika PKB dla Polski oraz dla: Belgii, Czech, Estonii, Francji, Niemiec oraz Bułgarii, gdzie uwagę zwrócono na wrażliwość fazy na wartość parametru wygładzającego. W analizie tej rozważono okres o raczej niskiej aktywności gospodarczej (bez wyraźnych oznak poprawy lub pogorszenia koniunktury).

2. KONCEPCJA ZEGARA CYKLU KONIUNKTURALNEGO

Zegar cyklu ilustruje cztery możliwe stany aktywności gospodarczej, zgodnie z przyjętymi w literaturze czterema fazami cyklu koniunkturalnego, to jest fazą ekspansji, spowolnienia, recesji oraz ożywienia. Naniesione na układzie współrzędnych punkty zegara, które opisują wartości poziomu cyklu odchyleń od trendu, oraz jego przyrostów wartości, tworzą trajektorię ruchu w kierunku przeciwnym do ruchu wskazówek zegara. Niech dla pewnego zaobserwowanego szeregu czasowego niech $\{C_t : t \in \mathsf{Z}\}$ będzie wyodrębnionym cyklem odchyleń od długookresowego trendu. Zegar cyklu koniunkturalnego jest zbiorem punktów w R^2 postaci $\{(C_t - C_{t-1}, C_t) : t \in \mathsf{Z}\}$.

Przejście rozważanej trajektorii przez pierwszą ćwiartkę układu współrzędnych wskazuje na okres poprawy koniunktury, z dodatnim wyhamowywanym tempem zmian cyklu odchyleń. Zbliżanie się trajektorii do drugiej ćwiartki układu współrzędnych wskazuje zaś na coraz słabszą dynamikę wzrostu aktywności. Prowadzi to do przejścia przez górny punkt zwrotny cyklu odchyleń i wejście trajektorii do drugiej ćwiartki układu współrzędnych. Następuje wtedy pogorszenie koniunktury, przy ujemnej stopie wzrostu cyklu odchyleń. Trzecia ćwiartka to kontynuacja okresu pogarszania się koniunktury, przy rosnącej stopie wzrostu cyklu odchyleń, jednak w dalszym ciągu ujemnej. Przejście z trzeciej do czwartej ćwiartki układu współrzędnych oznacza przejście przez tak zwany dolny punkt zwrotny cyklu odchyleń. W czwartej ćwiartce układu współrzędnych mamy do czynienia z okresem poprawy koniunktury, z dodatnią i rosnącą stopą wzrostu z cyklu odchyleń.

W niniejszym artykule przedstawiono również podejście alternatywne, w którym, przy prezentacji pozycji cyklicznej na zegarze, wzięto pod uwagę także element trendu zjawiska. Przedstawiono zatem na osi poziomej pierwsze różnice łącznie trendu i cyklu odchyleń, zaś na osi pionowej – wartości cyklu odchyleń. Wariant ten uwzględnia zatem (na osi poziomej) zmiany nie tylko wahań cyklicznych, lecz łączną dynamikę trendu i wahań cyklicznych. W konsekwencji punkty zegara w drugim wariancie są przesunięte w prawo w przypadku występowania trendu rosnącego oraz odpowiednio w lewo w stosunku do ścieżki pierwszego wariantu w przypadku obecności trendu odpowiednio malejącego. Pozycję cykliczną w przypadku obydwu wariantów wygodnie jest przedstawiać w kategoriach procentowych. Wariant 1 – klasyczny – na osi poziomej przedstawia przybliżone wartości procentowych miesięcznych zmian komponentu cyklicznego (cyklu odchyleń), czyli wielkości danej zmiennej, z pominięciem wahań sezonowych oraz trendu. W przypadku wariantu 2, na osi poziomej zaznaczono (przybliżone) procentowe zmiany miesięczne wielkości danej zmiennej, z pominięciem wahań sezonowych. Oś pionowa to (przybliżone) procentowe odchylenia wielkości danej zmiennej od linii trendu w danej chwili czasu.

Analiza zmian w koniunkturze może bazować na łącznym modelowaniu wahań cyklicznych, sezonowych i ogólnej tendencji rozwojowej obserwowanego makroekonomicznego szeregu czasowego. Takie podejście stosowali Luginbuhl, de Vos (2003) sugerując, że „*It is generally acknowledged that the growth rate of output, the seasonal pattern, and the business cycle are best estimated simultaneously.*" Jednak powszechnie przyjęta w analizach wahań cyklicznych procedura polega na wyodrębnieniu w pierwszym kroku wahań utożsamianych z wahaniami aktywności gospodarczej, poprzez eliminację wahań sezonowych metodami parametrycznymi. W drugim kroku stosuje się metody filtracji filtrami pasmowo przepustowymi. Takie podejście nie jest spójne ze względu na zastosowanie w pierwszym kroku podejścia parametrycznego, zaś w drugim kroku (w sposób niezależny od estymacji w kroku pierwszym) podejścia alternatywnego. W następnej części zaprezentowano propozycję spójnej konstrukcji zegara cyklu koniunkturalnego w podejściu nieparametrycznym.

3. ZASTOSOWANIE UJĘCIA NIEPARAMETRYCZNEGO W KONSTRUKCJI ZEGARA CYKLU KONIUNKTURALNEGO

W celu zdefiniowania nieparametrycznego zegara cyklu koniunkturalnego należy wprowadzić równanie nieparametrycznego modelu opisującego dynamikę jednowymiarowego procesu makroekonomicznego. Równanie to zaczerpnięto z pracy Lenart, Pipień (2013), w której autorzy zaproponowali niestandardową metodę wnioskowania statystycznego o naturze wahań cyklicznych, sezonowych i trendu z możliwością interakcji tych komponentów. Niech $\{P_t : t \in \mathsf{Z}\}$ będzie logarytmem naturalnym obserwowanego procesu makroekonomicznego. W równaniu modelu przyjmujemy, iż dla średniej procesu $\{P_t : t \in \mathsf{Z}\}$ (bezwarunkowej wartości oczekiwanej) zachodzi:

$$\mu_P(t) = E(P_t) = f(t,\beta) + g(t),$$

gdzie $g(t)$ jest funkcją prawie okresową postaci:

$$g(t) = \sum_{\psi \in \Psi_P} m_P(\psi) e^{i\psi t}, \tag{1}$$

zaś $f(t,\beta)$ jest wielomianem stopnia p o zerowym wyrazie wolnym; $\beta \in \mathsf{R}^p$. Nieznany zbiór $\Psi_P = \{\psi : m(\psi) \neq 0\} \subset [0,2\pi)$ można w naturalny sposób przedstawić jako:

$$\Psi_P = \Psi_{P,1} \cup \Psi_{P,2} \cup \Psi_{P,3}, \tag{2}$$

gdzie $\Psi_{P,1} \cap (0, \frac{2\pi}{1.5T}) = \Psi_{P,1}$, dla T oznaczającego liczbę obserwacji w roku dla procesu P_t (dane miesięczne lub kwartalne). Zbiór $\Psi_{P,1}$ zawiera zatem częstotliwości odpowiadające wahaniom dłuższym niż 1,5 roku. Zbiór $\Psi_{P,2}$ zawiera częstotliwości odpowiadające wahaniom sezonowym, tzn. $\Psi_{P,2} \subset \{2k\pi/12 : k = 0,1,\ldots,T-1\}$. Zbiór $\Psi_{P,3}$ jest zbiorem pozostałych częstotliwości, które nie są obiektem zainteresowania w tym artykule (np. częstotliwości korespondujące do efektu dni roboczych; patrz Ladiray, 2012). W celu wyodrębnienia wahań utożsamianych z wahaniami koniunkturalnymi zastosowano nieparametryczne podejście bazujące na filtracji. W celu osłabienia efektu wahań sezonowych zastosowano filtr $L_{2\times T}(B)$ scentrowanej średniej ruchomej 2xTMA. Działając tym operatorem otrzymujemy szereg czasowy $\{Y_t : t \in \mathsf{Z}\}$ dla którego:

$$Y_t = L_{2\times T}(B) P_t,$$

gdzie

$$L_{2\times T}(B) = (B^{-T/2} + 2B^{-T/2+1} + \ldots + 2B^{-1} + 2 + 2B + \ldots + 2B^{T/2-1} + B^{T/2})/(2T),$$

zaś $B^k P_t = P_{t-k}$ dla dowolnych całkowitoliczbowych wartości czasu t i przesunięcia k. Jak wykazano w pracy Lenart, Pipień (2013) zbiór $\Psi_{P,1}$ jest niezmienniczy ze względu na zastosowanie filtra $L_{2\times T}(B)$, zaś zbiór $\Psi_{P,2}$ jest co najwyżej zbiorem jednoelementowym zawierającym częstotliwość równą zero. Wynika to z tego, iż funkcja bezwarunkowej wartości oczekiwanej szeregu czasowego $\{Y_t : t \in \mathsf{Z}\}$ ma postać

$$\mu_Y(t) = E(Y_t) = \underbrace{\tilde{\beta}_0 + \tilde{\beta}_1 t + \ldots + \tilde{\beta}_p t^p}_{\tilde{f}(t,\tilde{\beta})} + \sum_{\psi \in \Psi_Y} m_Y(\psi) e^{i\psi t}, \tag{3}$$

gdzie $\Psi_Y \cap \{2k\pi/T : k = 1,2,\ldots,T-1\} = \varnothing$ oraz $\Psi_Y = \Psi_P \setminus \{2k\pi/12 : k = 1,2,\ldots,T-1\}$. Zatem szereg czasowy $\{Y_t : t \in \mathsf{Z}\}$ zawiera te same częstotliwości w zbiorze Ψ_Y co szereg czasowy $\{P_t : t \in \mathsf{Z}\}$ w zbiorze Ψ_P, po odjęciu częstotliwości identyfikowanych z wahaniami sezonowymi. Dodatkowo, dla współczynników $m_P(\psi)$ oraz $m_Y(\psi)$ zachodzi zależność

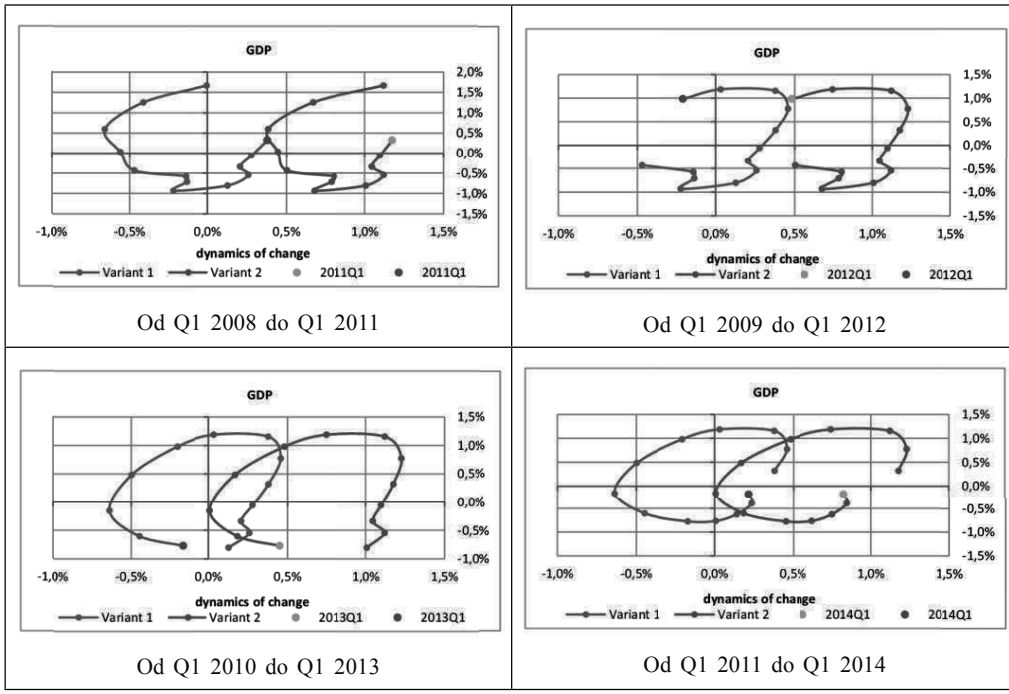
$$m_Y(\psi) = L_{2\times T}(e^{-i\psi}) m_P(\psi). \tag{4}$$

Dla współczynników $\tilde{\beta}_k$ oraz β_k otrzymujemy $\tilde{\beta}_k = \beta_k$, dla $k = p$ oraz $k = p - 1$, co oznacza, że dla $p = 0$ oraz $p = 1$ współczynniki odpowiednich wielomianów nie ulegają zmianie. W przypadku zaś, gdy $p = 2$ może się zmieniać jedynie wyraz wolny wielomianu.

W kolejnym kroku wyodrębniony zostanie komponent cykliczny przy użyciu filtra HP $L_{HP(\lambda)(B)}$ z parametrem wygładzającym $\lambda > 0$. Jeśli wielomian $f(t,\beta)$ jest wielomianem stopnia co najwyżej czwartego wtedy dla wyodrębnionego procesu wahań cyklicznych $C_t = L_{HP(\lambda)}(B) Y_t$ mamy $E(C_t) = E(L_{HP(\lambda)}(B) Y_t) = L_{HP(\lambda)}(B) E(Y_t) =$ $= \sum_{\psi \in \Psi_C} m_C(\psi) e^{i\psi t}$, gdzie $\Psi_C \setminus \{0\} = \Psi_Y \setminus \{0\}$, co oznacza iż zbiory te różnią się co najwyżej jedną częstotliwością równą zero oraz $m_C(\psi) = L_{HP(\lambda)}(e^{-i\psi}) m_Y(\psi) =$ $= L_{HP(\lambda)}(e^{-i\psi}) L_{2\times TMA}(e^{-i\psi}) m_Y(\psi)$ (patrz twierdzenie 1 w Dodatku). Zbiór $\Psi_{P,1}$ jest zatem niezmienniczy ze względu na zastosowane nieparametrycznej metody filtracji. Oznacza to, że punkty zegara mające współrzędne $\{(C_t - C_{t-1}, C_t) : t \in \mathsf{Z}\}$ mogą być interpretowane jako proces stochastyczny o prawie okresowej dwuwymiarowej funkcji wartości oczekiwanej.

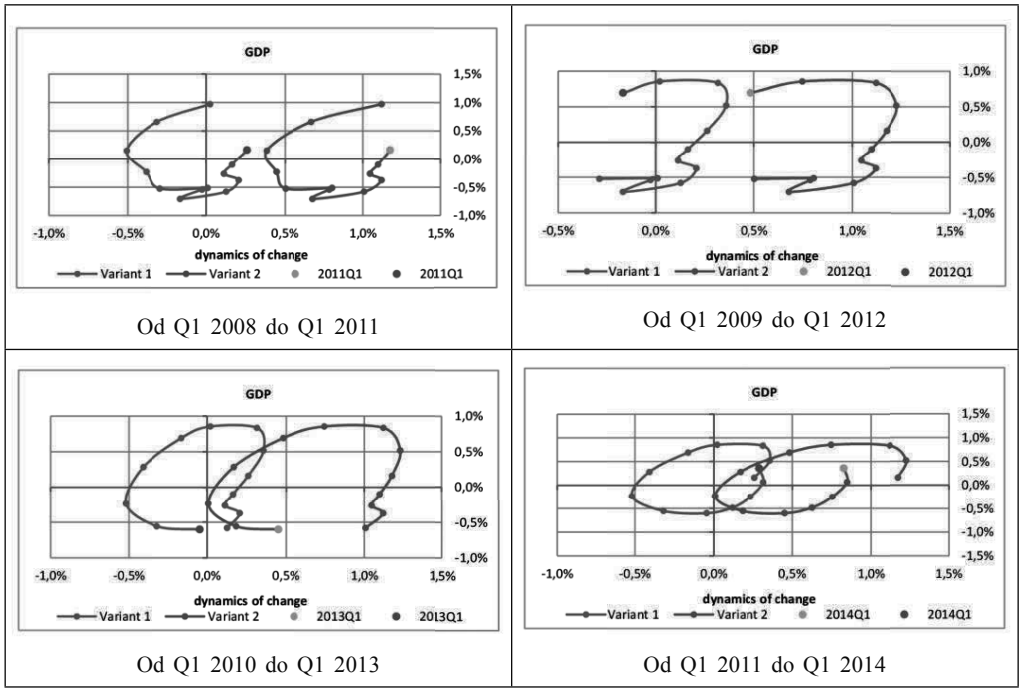
Rozważmy przykład, który zilustruje zasady działania powyższej konstrukcji. Na rysunkach 1 i 2 przestawiono zegary cyklu koniunkturalnego uzyskanego na podstawie szeregu czasowego PKB w Polsce (GDP and main components, Chain linked volumes, index 2005 = 100, NSA) dla dwóch różnych wartości parametru wygładzającego λ, odpowiadających granicznemu okresowi wahań odpowiednio 4 i pół roku oraz 8 lat. Dla każdego parametru λ przedstawiono cztery zegary (w ruchomym oknie 13 obserwacji w okresie od pierwszego kwartału 2008 r. do pierwszego kwartału 2014 r.).

Trajektorie umieszczone na zegarach należy odczytywać jako opis ruchu w kierunku przeciwnym do ruchu wskazówek zegara. Punkty zegara w drugim wariancie są wyraźnie przesunięte w prawą stronę i praktycznie dla wszystkich okien znajdują się po prawej stronie osi OY. Wskazuje to na dodatnie co do wartości miesięczne przyrosty analizowanego wskaźnika (z pominięciem wahań sezonowych). Wszystkie punkty zegara w wariancie 1 znajdujące się w trzeciej ćwiartce układu współrzędnych wskazują na występowanie fazy spowolnienia gospodarczego – jednak z dodatnimi wartościami kwartalnych zmian (z pominięciem wahań sezonowych – patrz wariant 2).



Rysunek 1. Zegary cykli koniunkturalnych uzyskane na podstawie szeregu czasowego realnego PKB Polski – parametr wygładzania λ odpowiada cyklowi 8 lat

Źródło: opracowanie własne.



Od Q1 2008 do Q1 2011

Od Q1 2009 do Q1 2012

Od Q1 2010 do Q1 2013

Od Q1 2011 do Q1 2014

Rysunek 2. Zegary cykli koniunkturalnych uzyskane na podstawie szeregu czasowego realnego PKB Polski – parametr wygładzania λ odpowiada cyklowi 4 i pół roku

Źródło: opracowanie własne.

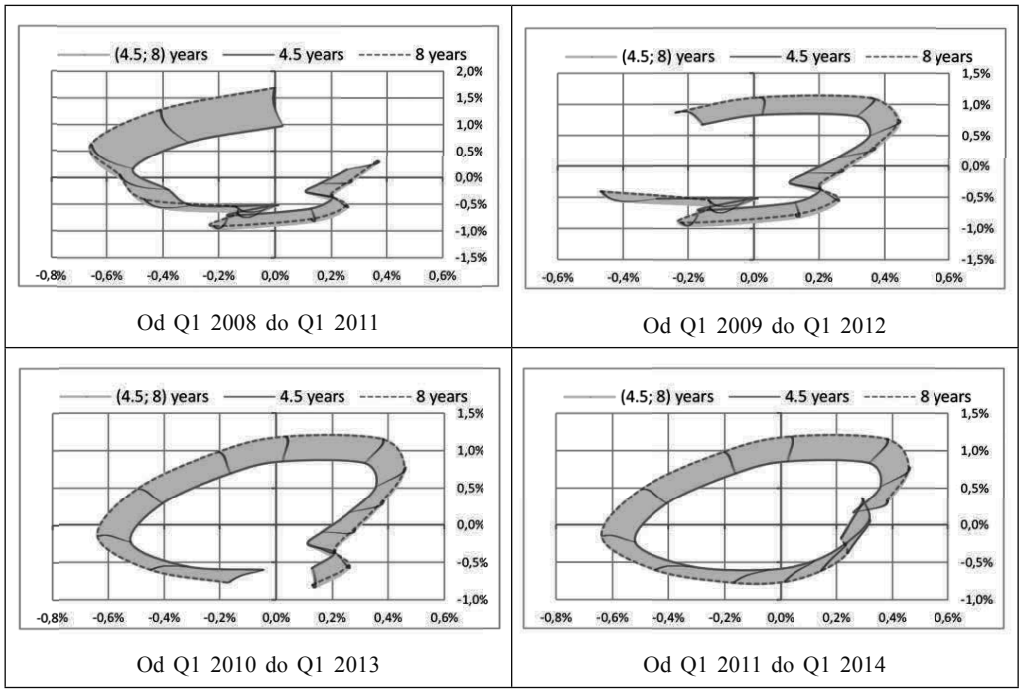
Trajektorie wyznaczone przez punkty zegara zależą od wartości parametru wygładzającego metody HP. W przypadku czwartego z rozpatrywanych okresów (to jest dla okresu czasu od pierwszego kwartału 2011 do pierwszego kwartału 2014) ostatni punkt zegara znajduje się w pierwszej ćwiartce układu współrzędnych, jeśli przyjmie się wartość parametru wygładzającego λ odpowiadającego cyklowi 4 i półrocznemu (rysunek 1). Natomiast dla wartości λ odpowiadającej cyklowi ośmioletniemu (rysunek 2) ostatni punkt zegara pozostaje nadal w czwartej ćwiartce. Oznacza to, że w pewnych szczególnych przypadkach badacz może mieć poważne kłopoty z precyzyjnym odczytaniem fazy cyklu w ramach analizowanej metody. Skoro w prawidłowym określeniu przebiegu trajektorii na zegarze kluczową rolę odgrywa wybór parametru wygładzającego λ, należy przyjrzeć się jego roli w stosowanym filtrze.

Literatura nie podaje jednoznacznej odpowiedzi jaką wartość parametru wygładzania λ należy wybrać. Ogólne względy metodologiczne sugerują aby wartości parametru λ były różne dla danych o różnej częstotliwości obserwacji. Zmiana parametru λ wpływa na przebieg wyodrębnionej linii trendu w ten sposób, że im większa wartość parametru λ, tym uzyskana linia trendu jest gładsza, a przez to wyodrębnione wahania (będące różnicą pomiędzy danymi a wartościami z linii trendu) zawierają cykle

o większej długości. Algorytm doboru parametru wygładzającego λ w zależności od długości cykli będących obiektem zainteresowania zaprezentowano w pracy Maravall, del Rio (2001), gdzie przytoczono formułę na wartość parametru λ jako funkcję częstotliwości ω_0 postaci:

$$\lambda = \frac{1}{4(1-\cos(\omega_0))^2}. \tag{5}$$

Wartość ω_0 można interpretować jako dolną granicę częstotliwości będących obiektem zainteresowania. Przy ustalonej wartości ω_0, wyliczona na podstawie (5) wartość parametru λ może być interpretowana jako granica, dla której, po zastosowaniu filtra HP, wzmocnione zostaną wahania o częstotliwościach powyżej wartości ω_0. Osłabieniu zaś podlegać będą wahania o częstotliwościach poniżej wartości ω_0. Ta interpretacja bazuje na wykazaniu, że filtr HP można otrzymać jako szczególny przypadek filtra Butterwortha, Gómez (1999), Gómez (2001).



Rysunek 3. Wstęgowe zegary cyklu koniunkturalnego uzyskane na podstawie szeregu czasowego realnego PKB Polski – graniczne wartości parametru wygładzania λ_{min} i λ_{max} odpowiadają okresom 4 i pół roku oraz 8 lat

Źródło: opracowanie własne.

Rezultaty analizy wrażliwości przebiegu trajektorii cyklu na zegarze można przedstawić w postaci całego pasma punktów, które będziemy nazywać zegarem wstęgowym cyklu koniunkturalnego. W proponowanym podejściu na zegarze nie jest prezentowana pojedyncza trajektoria a cała wiązka trajektorii, uzyskanych na podstawie wyboru różnych wartości parametru wygładzania $\lambda \in [\lambda_{\min}, \lambda_{\max}]$, dla pewnych granicznych wartości $\lambda_{\min}$ i $\lambda_{\max}$. Odpowiada to analizie filtru HP w przedziale odpowiadających parametrom λ częstości granicznych $[\omega_{\min}, \omega_{\max}]$. Na rysunku 3 przedstawiono pasmo skonstruowane poprzez przyjęcie wartości parametru wygładzania λ w ten sposób, że przedział częstości granicznych odpowiada wahaniom nie krótszym niż 4 i pół roku i nie dłuższym niż 8 lat. Wybór częstotliwości granicznych jest w pewnym sensie arbitralny i ma w głównej mierze cel ilustracyjny. Dotychczasowe analizy autorów niniejszego opracowania w zakresie badań nad naturą cyklu koniunkturalnego w Polsce pokazują, że cykl ten może być opisany jako złożenie kilku komponentów cyklicznych o długości odpowiednio 2 lata, około czterech lat i 8 lat; por. Lenart, Pipień (2013), Lenart i inni (2016). Ten okołoczteroletni komponent jest interpretowany jako cykl produkcji.

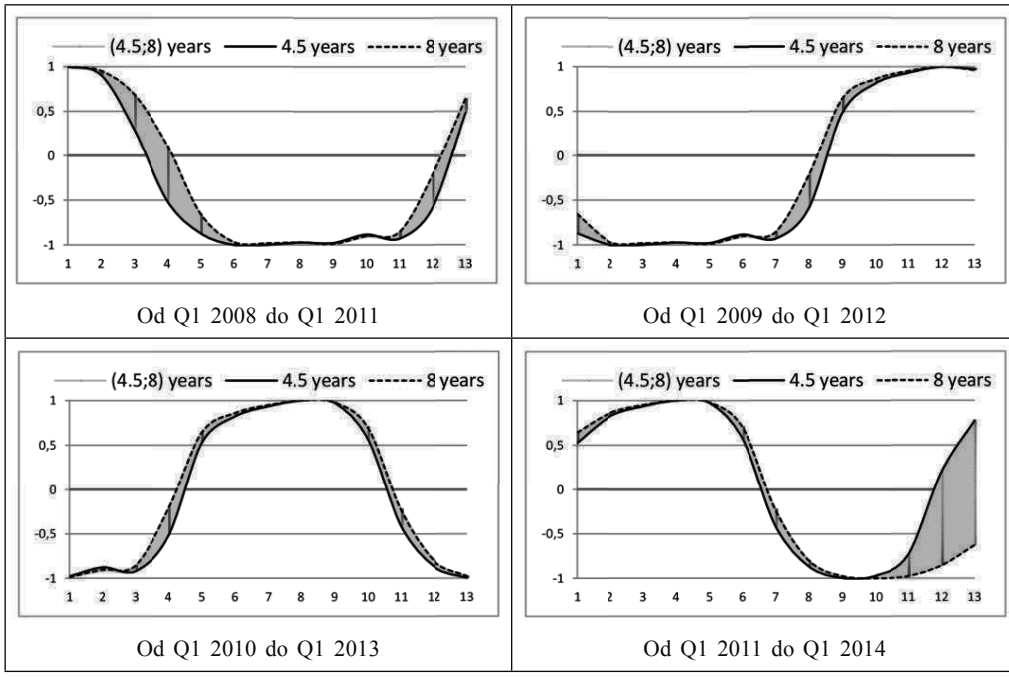
Widoczne na rysunku 3 pasma mają charakter regularny. Nie w każdym jednak okresie czasu ścieżki odpowiadające wahaniom granicznym 4 i pół roku oraz 8 lat wyznaczają obwiednię prezentowanej wstęgi. Dyskutowana forma prezentacji cykliczności na zegarze umożliwia przedstawienie w jakościowy sposób niepewności co do kształtowania się faz cyklu jak i ich czasów trwania. Wstęgowe zegary cyklu informują, iż w pewnych przypadkach analizy mogą nie dawać jednoznacznych konkluzji co do aktualnej fazy cyklu, zaś szerokość pasma może być interpretowana jako miara stopnia niepewności co do właściwego jej określenia.

4. WYODRĘBNIENIE FAZY I AMPLITUDY

Podstawowe charakterystyki cyklu koniunkturalnego to faza oraz amplituda. Problemem wartym podjęcia jest wypracowanie metody ilustracji tych dwóch charakterystyk w sposób odseparowany. Wyodrębnienie fazy cyklu z pominięciem jego amplitudy może bowiem dostarczyć bardziej precyzyjnego określenia stanu koniunktury, bez wnikania w szczegóły dotyczące głębokości jej wahań. W tym celu wprowadzimy pojęcie wykresu wstęgowego fazy. Niech będzie ustalona wartość parametru $\lambda \in [\lambda_{\min}, \lambda_{\max}]$ dla wstęgowego zegara cyklu. Sinus kąta jaki tworzy prosta przechodząca przez początek układu współrzędnych i dany punkt zegara, tj.

$$Z_{t,\lambda} = \frac{C_{t,\lambda}}{\sqrt{(C_{t,\lambda} - C_{t-1,\lambda})^2 + C_{t,\lambda}^2}} \tag{6}$$

będziemy nazywać fazą cyklu koniunkturalnego. Zbiór punktów $\{Z_{t,\lambda} : \lambda \in [\lambda_{\min}, \lambda_{\max}]\}$, gdzie $t \in \mathbb{Z}$ będziemy nazywać wykresem wstęgowym fazy.

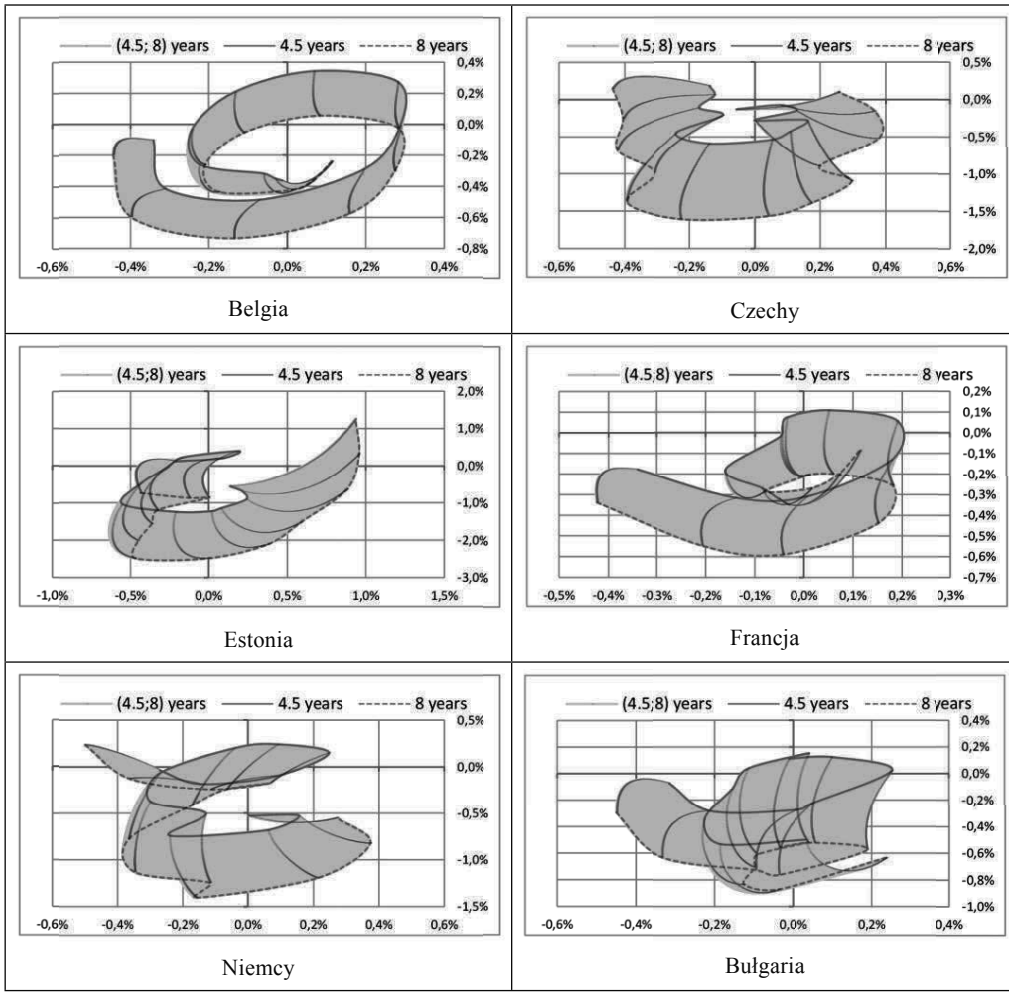


Rysunek 4. Wstęgowe wykresy fazy cyklu uzyskane na podstawie szeregu czasowego realnego PKB Polski – graniczne wartości parametru wygładzania λ_{min} i λ_{max} odpowiadają okresom 4 i pół roku oraz 8 lat

Źródło: opracowanie własne.

Rysunek 4 przedstawia wykresy przebiegu faz cyklu, otrzymanych zgodnie z formułą (6). Wiązka wykresów fazy jest uzyskana, podobnie jak na rysunku 3, poprzez przyjęcie parametrów wygładzania odpowiadających okresowi cyklu od 4 i pół roku do 8 lat. Wykresy przedstawiają wartości funkcji sinus obliczanych dla kolejnych punktów zegara cyklu w ruchomym oknie obserwacji analogicznym do ilustracji zawartych na rysunkach 1 i 2. Przebiegi wykresów fazy nie zmieniają się zasadniczo wraz ze zmianami parametru wygładzania λ. Wstęgowy wykres przebiegu faz jest bardziej skoncentrowany na skali wartości niż analogiczny wykres wstęgowy zegarów; por. rysunek 3. Zatem odizolowanie informacji o samej fazie cyklu od amplitudy wahań powoduje, że określenie stanu koniunktury wydaje się łatwiejsze. Jedynie w przypadku analiz cykliczności w okresie od 1 kwartału 2011 do 1 kwartału 2014 znacznej niepewności podlega określenie fazy cyklu pod koniec próby, to jest w przypadku 4 kwartału 2013 i 1 kwartału 2014; por. rysunek 4, wykres (d). Ale i w tym przypadku odizolowanie fazy od amplitudy wahań, zgodnie z formułą (6), daje obraz bardziej czytelny od informacji zawartej na zegarze koniunktury. W przypadku 4 kwartału 2013 (obserwacja 12 na wykresie (d)) większość przebiegów, uzyskanych na podstawie wyboru poszczególnej wartości λ, wskazuje na fazę ożywienia po recesji, natomiast w znacznie mniejszej liczbie przypadków okres ten może być utożsamiany z fazą

ekspansji. W przypadku 1 kwartału 2014 (obserwacja 13 na wykresie (d)) określenie fazy cyklu jest trudne. Niemal w połowie przypadków wartości parametru λ punkt określający pozycję cykliczną wskazuje na fazę ożywienia a w drugiej połowie wartości na fazę ekspansji. Analogiczną sytuację, choć nie tak wyraźnie zaakcentowaną, ilustruje wykres (a) na rysunku 4. W przypadku 4 kwartału 2008 (obserwacja 4) dla większości przyjętych wartości parametru λ pozycja cykliczna w tym okresie powinna być określona jako recesja. Niektóre zaś wartości λ lokalizują trajektorię w obszarze wartości dodatnich, co może prowadzić do wniosku, że w badanym okresie czasu mamy do czynienia jeszcze z wyhamowaniem a nie z recesją.



Rysunek 5. Wstęgowe zegary cyklu koniunkturalnego uzyskane na podstawie szeregów czasowych realnego PKB wybranych krajów europejskich – graniczne wartości parametru wygładzania λ_{min} i λ_{max} odpowiadają okresom 4 i pół roku oraz 8 lat

Źródło: opracowanie własne.

Określenie aktualnej fazy cyklu na podstawie analizy zegarów cykli koniunkturalnych może nie zawsze być łatwe i precyzyjne. Na rysunku 5 przedstawiono kolejne przykłady zegarów cykli o silnie zróżnicowanej czytelności. Cykliczność koniunktury określono w okresie od pierwszego kwartału 2003 do pierwszego kwartału 2006 w przypadku Belgii, Czech, Estonii, Francji, Niemiec i Bułgarii. Okres ten był dla większości gospodarek europejskich okresem o stosunkowo niskiej aktywności gospodarczej, w odniesieniu do innych okresów, bez oznak wyraźnego ożywienia bądź recesji. Taki stan gospodarki, polegający na wybiciu z naturalnego rytmu faz sprawia, iż niepewność co do określenia perspektyw rozwojowych jest duża. Na zegarze cyklu stan ten jest identyfikowany w przypadku, gdy punkty zegara znajdują się blisko początku układu współrzędnych. Na przykład w pierwszym i drugim kwartale 2003 roku dla Estonii wartości cyklu odchyleń oraz przyrosty wartości są bliskie zera. Dodatkowo, w przypadku przyjęcia różnych wartości parametru λ punkty te znajdują się w różnych ćwiartkach układu współrzędnych. Podobnie, precyzyjne określenie fazy cyklu w przypadku Bułgarii może przysparzać sporo kłopotów.

Oznacza to, że dla pełnego obrazu konieczne jest zdefiniowanie odrębnej fazy cyklu, którą będziemy nazywać fazą neutralną. Niech $\{C_{t,\lambda} : t \in \mathsf{Z}\}$ będzie wyodrębnionym cyklem odchyleń dla pewnego λ. Wtedy będziemy mówić o okresie fazy neutralnej jeśli jednocześnie wartości cyklu odchyleń $C_{t,\lambda}$ oraz przyrosty wartości tego cyklu $C_{t,\lambda} - C_{t-1,\lambda}$ będą bliskie zera, tj.:

$$\frac{(C_{t,\lambda})^2}{a^2} + \frac{(C_{t,\lambda} - C_{t-1,\lambda})^2}{b^2} < 1, \tag{7}$$

gdzie $a,b > 0$ są odpowiednio progami dla wartości cyklu odchyleń oraz różnic $C_{t,\lambda} - C_{t-1,\lambda}$. Niskie wartości $C_{t,1}$ świadczą o rozwoju w tempie bliskim potencjału zaś niskie wartości $C_{t,\lambda} - C_{t-1,\lambda}$ świadczą o tym, iż odstępstwa od potencjału nie są duże.

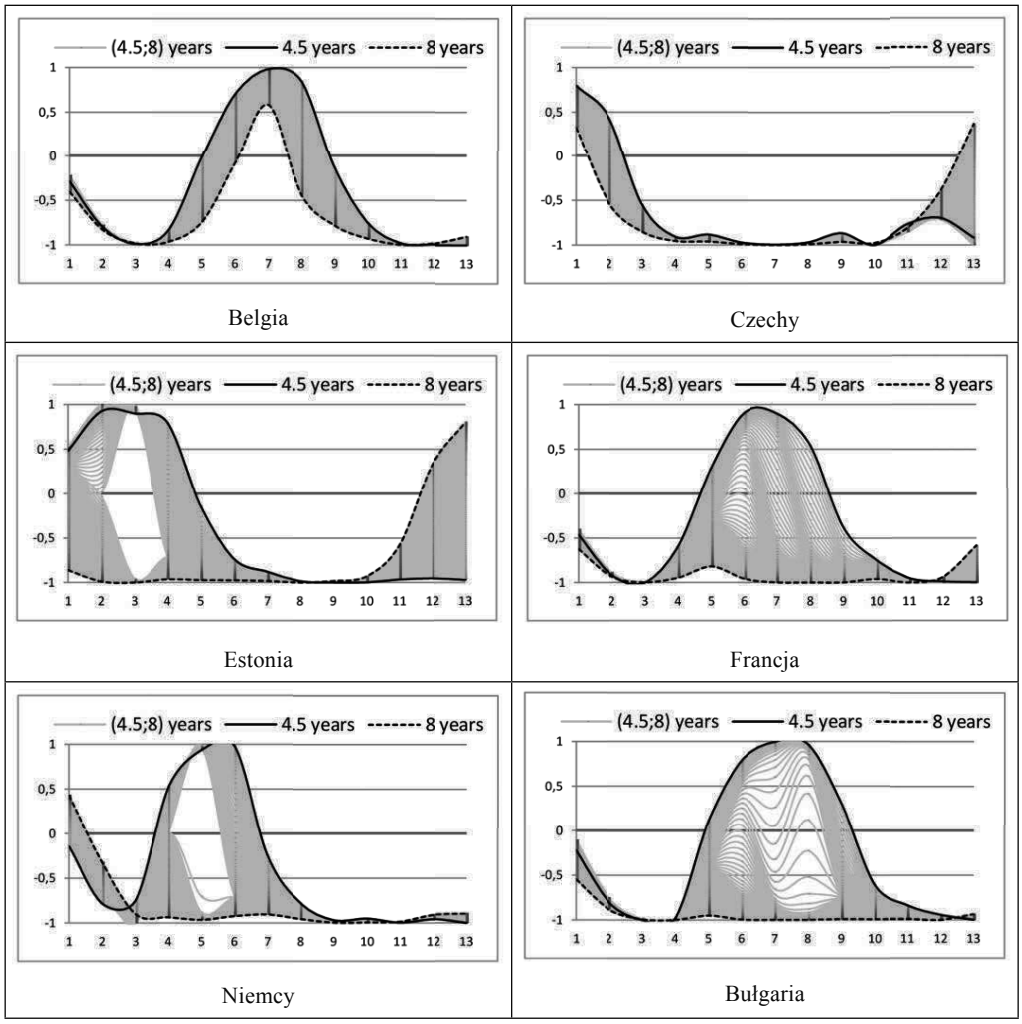
Dobór parametrów a oraz b powinien zależeć od własności procesu cyklicznego $C_{t,\lambda}$, jednak nie jest przedmiotem naszych rozważań. Nierówność

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} < 1,$$

spełniają punkty (x,y) które leżą wewnątrz elipsy o półosiach a oraz b.

Rezultaty odseparowania informacji o fazie cyklu od amplitudy, uzyskane w przypadku analizowanych sześciu krajów, zawarto na rysunku 6. Podobnie jak na rysunku 4, przedmiotem analiz jest przebieg faz cyklu otrzymanych według formuły (6). Wiązkę wykresów uzyskano poprzez przyjęcie parametrów wygładzania odpowiadających okresowi cyklu od 4 i pół roku do 8 lat. Odizolowanie informacji o samej fazie cyklu od amplitudy wahań nie powoduje tak znaczącej poprawy w precyzyjnym określeniu stanu koniunktury analizowanych krajów, jak w przypadku Polski; por. rysunek 4. Szczególnie, określenie faz ożywienia i ekspansji, które występują w różnych okre-

sach czasu dla poszczególnych krajów, jest obarczone dużą niepewnością. Dla Belgii i Czech fazy spowolnienia i recesji są określone precyzyjnie i przebieg komponentu cyklicznego, z pominięciem informacji o amplitudzie, prawie nie zmienia się w tych przypadkach, wraz ze zmianami wartości parametru wygładzania. W przypadku Estonii, Francji, Niemiec i Bułgarii określenie fazy cyklu w większości kwartałów jest niemożliwe. Komponent cykliczny z pominięciem informacji o amplitudzie jest bardzo wrażliwy na zmiany wartości parametru wygładzania i może wskazywać zarówno na ożywienie, jaki i na recesję.



Rysunek 6. Wstęgowe wykresy fazy cyklu uzyskane na podstawie szeregu czasowego realnego PKB wybranych krajów europejskich – graniczne wartości parametru wygładzania λ_{min} i λ_{max} odpowiadają okresom 4 i pół roku oraz 8 lat

Źródło: opracowanie własne.

5. PODSUMOWANIE

W pracy przedstawiono konstrukcję wstęgowego zegara cyklu koniunkturalnego (bazującego na jednowymiarowym realnym wskaźniku makroekonomicznym) z uwzględnieniem przedziału dla parametru wygładzającego metody filtracji. Przyjęcie takiej koncepcji zegara powoduje, iż interpretacji podlega całe spektrum punktów zegara a nie tylko pojedyncza ścieżka. W celu trafniejszego opisu aktualnej pozycji cyklicznej na zegarze cyklu przedstawiono prosty sposób odseparowania fazy cyklu od amplitudy. Również ta konstrukcja uwzględnia przedział parametrów wygładzających metody filtracji HP.

Przykłady analizy danych empirycznych w pracy mają charakter ilustracyjny i ich analiza koncentrowała się bardziej na wskazaniu użyteczności skonstruowanych narzędzi niż na weryfikacji ewentualnych hipotez badawczych dla analizowanych gospodarek i okresów. Praca wskazuje, iż skonstruowane narzędzia oparte na realnych wskaźnikach makroekonomicznych mogą być pomocne w badaniach nad naturą cykli koniunkturalnych.

Wstęgowe zegary cyklu informują, iż w pewnych przypadkach analizy mogą nie dawać jednoznacznych konkluzji co do aktualnej fazy cyklu. Szerokość pasma może być interpretowana jako miara stopnia niepewności co do właściwego jej określenia. Odizolowanie informacji o samej fazie cyklu od amplitudy wahań powoduje, że określenie stanu koniunktury wydaje się łatwiejsze. Dodatkowo takie podejście skłania do określenia alternatywnej fazy cyklu, którą autorzy proponują nazywać neutralną.

LITERATURA

Abberger K., Nierhaus W., (2010), IFO Business Cycle Clock: Circular Correlation with the Real GDP, CESifo Working Paper Series No. 3179.

Gómez V., (1999), Three Equivalent Methods for Filtering Finite Nonstationary Time Series, *Journal of Business and Economic Statistics*, 17 (1), 109–117.

Gómez V., (2001), The Use of Butterworth Filters for Trend and Cycle Estimation in Economic Time Series, *Journal of Business and Economic Statistics*, 19 (3), 365–373.

Hodrick R., Prescott E., (1997), Postwar U.S. Business Cycles: An Empirical Investigation, *Journal of Money, Credit and Banking*, 29, 1–16.

Ladiray D., (2012), Theoretical and Real Trading-Day Frequencies, w: Bell W. R., Holan S. H., McErloy T. S., (red.), Economic Time Series: Modeling and Seasonality, 255–279.

Lenart Ł., Pipień M., (2013), Almost Periodically Correlated Time Series in Business Fluctuations Analysis, *Acta Physica Polonica A*, 123 (3), 567–583.

Lenart Ł., Mazur B., Pipień M., (2016), Statistical Analysis of Business Cycle Fluctuations in Poland Before and After the Crisis, *Equilibrium. Quarterly Journal of Economics and Economic Policy* (in press).

Luginbuhl R., de Vos A., (2003), Seasonality and Markov Switching in an Unobserved Component Time Series Model, *Empirical Economics*, 28, 365–386.

DODATEK

Twierdzenie 1. Jeśli $f(t)$ jest wielomianem stopnia co najmniej czwartego, oraz $T(L) = \dfrac{1}{\lambda L^{-2} - 4\lambda L^{-1} + 6\lambda + 1 - 4\lambda L^{1} + \lambda L^{2}}$ jest operatorem trendu w filtrze HP, takim, że $L^k f(t) = f(t - k)$ dla każdego $k, t \in \mathsf{Z}$, wtedy $T(L) f(t)$ i $f(t)$ różnią się co najwyżej o stałą.

Dowód. Niech $f(t)$ będzie wielomianem rzędu co najwyżej cztery, $f(t) = at^4 + bt^3 + ct^2 + dt + e$. Wtedy wielomian $w(t) = T(L) f(t)$ jest równoważny.

$$f(t) = \lambda w(t-2) - 4\lambda w(t-1) + (6\lambda + 1)w(t) - 4\lambda w(t+1) + \lambda w(t+2) \qquad (8)$$

Rozważmy reprezentację $w(t) = f(t) + z(t)$. Na jej podstawie otrzymujemy:

$$-24a\lambda = \lambda z(t-2) - 4\lambda z(t-1) + (6\lambda + 1)z(t) - 4\lambda z(t+1) + \lambda z(t+2). \qquad (9)$$

Powyższe równanie funkcyjne ma rozwiązanie postaci $z(t) = -24a\lambda$. Skoro $w(t)$ jest określony jednoznacznie, to $z(t)$ jest funkcją stałą.

KONCEPCJA WSTĘGOWEGO ZEGARA CYKLU KONIUNKTURALNEGO W UJĘCIU NIEPARAMETRYCZNYM

Streszczenie

W artykule omówiono propozycję uwzględnienia niepewności na zegarze cyklu koniunkturalnego. Stosowane podejście bazuje na reprezentacji wartości oczekiwanej realnych wskaźników makroekonomicznych jako sumy trendu i funkcji prawie okresowej w ramach równania nieparametrycznego. Ujmujemy w ten sposób łącznie dynamikę wahań koniunkturalnych, sezonowych, trendu i możliwej interakcji pomiędzy tymi komponentami. Poprzez zastosowanie nieparametrycznych metod filtracji, w celu eliminacji wahań sezonowych oraz trendu, uzyskano wartość pierwszego momentu punktów zegara jako poszczególne wartości funkcji prawie okresowej. Częstotliwości utożsamiane z wahaniami aktywności gospodarczej, jak również te, które charakteryzują dynamikę zegara cyklu, są niezmiennicze ze względu na stosowane metody filtracji.

Słowa kluczowe: zegar cyklu koniunkturalnego, metody filtracyjne, procesy prawie okresowo skorelowane

NONPARAMETRIC BAND BUSINESS CYCLE CLOCK

Abstract

We discuss representation of uncertainty in the business cycle clock. We propose approach utilising description of the unconditional mean of the process, applied for modelling dynamics of macroeconomic time series, as a trend component and almost period function in a non-parametric setting. We capture the dynamics over the business cycle, trend component and seasonal fluctuations and possible interactions between these features. A particular values of the almost periodic function are key for representation of the business cycle in a clock, expressing the dynamics according to phase diagram. The set of frequencies interpreted as a properties of the business fluctuations are invariant with respect to filtration methods applied in the procedure.

Keywords: business cycle clock, filtration, almost periodically correlated processes

DOI: 10.5604/01.3001.0014.1222

PRZEGLĄD STATYSTYCZNY
R. LXIII – ZESZYT 4 – 2016

HENRYK GURGUL[1], PAWEŁ ZAJĄC[2]

ZASTOSOWANIE MODELU PRZYSPIESZONEJ PORAŻKI DO ANALIZY PRZEŻYCIA MAŁOPOLSKICH PRZEDSIĘBIORSTW

1. WSTĘP

Powstawanie i likwidacja przedsiębiorstw, jako nieodłączne elementy mechanizmu rynkowego, decydują o obrazie współczesnej gospodarki. W odpowiedzi na zmieniające się potrzeby rynku powstają nowe podmioty gospodarcze oraz upadają przedsiębiorstwa, które nie są w stanie sprostać konkurencji. Nowe jednostki gospodarcze, wprowadzając innowacje, stają się motorem rozwoju gospodarczego, a jednocześnie są przyczyną upadku podmiotów, które nie nadążają za postępem. Likwidacja przedsiębiorstw pełni rolę katalizatora, oczyszcza rynek z nieefektywnych jednostek i tworzy miejsce dla nowych. Selekcja wśród przedsiębiorstw w dużej mierze kształtuje rozwój gospodarki regionu i całego kraju.

Ze względu na decydujące znaczenie kondycji przedsiębiorstw dla szeroko pojętego rozwoju gospodarczego regionu (poziom zatrudnienia, wzrost gospodarczy, konkurencyjność gospodarki) ścisłe określenie czynników oddziałujących na nie zarówno w krótkim, jak i długim okresie jest istotnym problemem stojącym przed badaczami. W publikacjach teoretycznych i empirycznych stosuje się podział czynników wpływających na powstawanie, rozwój i przeżywalność przedsiębiorstw na czynniki endogeniczne i egzogeniczne. Wśród czynników endogenicznych wyróżnia się najczęściej bodźce finansowe, które związane są z odpowiednim finansowaniem aktywów i utrzymaniem płynności finansowej, a także czynniki menedżerskie, związane z podejmowaniem trafnych decyzji oraz przyjmowaniem właściwych postaw przez zarządzającego przedsiębiorstwem menedżera. Wśród czynników endogenicznych można wymienić ponadto wiek przedsiębiorstwa, jego wielkość, rodzaj prowadzonej działalności gospodarczej i region, w którym firma operuje. Czynniki egzogeniczne podzielić można na czynniki branżowe oraz na czynniki ogólnoekonomiczne (w szczególności makroekonomiczne), dotyczące całego kraju. Wyróżnić tutaj można poziom konkurencji, koszty pracy, kurs walutowy, kapitał ludzki, zamożność społeczeństwa, wysokość inflacji,

[1] AGH Akademia Górniczo-Hutnicza, Wydział Zarządzania, Samodzielna Pracownia Zastosowań Matematyki w Ekonomii, Al. Mickiewicza 30, 30-059 Kraków, Polska, autor prowadzący korespondencję – e-mail: henryk.gurgul@gmail.com.

[2] AGH Akademia Górniczo-Hutnicza, Wydział Zarządzania, Samodzielna Pracownia Zastosowań Matematyki w Ekonomii, Al. Mickiewicza 30, 30-059 Kraków, Polska.

dostępność kredytów, poziom biurokracji, stabilność przepisów, wysokość podatków, postęp technologiczny. Nie bez znaczenia pozostaje również pomoc finansowa ze środków Funduszy Europejskich oraz pomoc instytucji wspierania biznesu w regionie. W rzeczywistości często zdarza się, że czynniki zewnętrzne stymulują te wewnętrzne, w równej mierze dotyczy to zarówno rozwoju jak i upadłości przedsiębiorstw.

Tematyką demografii przedsiębiorstw w Polsce zajmowały się m.in. Ptak-Chmielewska (2010) i Markowicz (2012). W swoich badaniach korzystały one z narzędzi analizy czasu trwania, w tym z tablic trwania, estymatora Kaplana-Meiera, modelu Coxa oraz modeli logitowych. Z modelem logitowym pracował również Pociecha (2012), przeanalizował on zalety i ograniczenia stosowalności modelu w badaniach bankructw w Polsce.

Nehrebecka, Dzik (2013) wykorzystując regresję logistyczną analizowały wpływ indywidualnych danych bilansowych oraz danych z raportu dochodów polskich spółek na prawdopodobieństwo bankructwa polskich przedsiębiorstw.

Przedmiotem niniejszego artykułu jest weryfikacja wpływu szeregu czynników mierzalnych na dynamikę populacji przedsiębiorstw w Małopolsce. Wykorzystanie metod analizy przeżycia, w tym w szczególności modelu przyspieszonej porażki, pozwoli sprecyzować, w jaki sposób czynniki związane z miejscem i rodzajem rozpoczęcia działalności gospodarczej wpływają na czas jej trwania.

2. PRZEGLĄD LITERATURY

Powody skłaniające ludzi do rozpoczynania działalności gospodarczej mogą być bardzo zróżnicowane. Według van Praag (1996) decyzja o założeniu firmy jest pochodną chęci i możliwości. Opisując „chęci" wskazuje ona bardziej szczegółowo na indywidualne preferencje uzależnione od posiadanych cech charakteru przedsiębiorcy. Chęci zależą również w znacznej mierze od występowania opcji alternatywnych i ich atrakcyjności w oczach potencjalnego przedsiębiorcy. Mówiąc o „możliwościach", badaczka wskazuje na czynniki mikroekonomiczne, np. dostępność kapitału początkowego, i makroekonomiczne, związane z ogólną koniunkturą gospodarczą.

W sposób naturalny koniunkturę gospodarczą w regionie można łączyć z rynkiem pracy. Już Knight (1921) twierdził, że człowiek ma trzy wyjścia, może być bezrobotny, może zostać pracownikiem najemnym, albo może się sam zatrudnić. W związku z tym decyzja o założeniu działalności gospodarczej wynika z prostej kalkulacji zysków i strat płynących z wyboru poszczególnych ścieżek. Można zatem na czynniki rozwoju przedsiębiorczości spojrzeć w ten sposób, że determinantą decyzji o uruchomieniu własnego przedsiębiorstwa jest zmniejszenie atrakcyjności alternatyw, czyli zredukowanie świadczeń przysługujących bezrobotnym lub spadek atrakcyjności pracy na stanowisku najemnym, wynikający np. ze wzrostu konkurencji spowodowanego wzrostem bezrobocia.

Decyzja o założeniu działalności gospodarczej jest silnie uzależniona od czynników geograficznych (Armington, Acs, 2002). Fritsch (1997) twierdzi, że powstanie

nawet połowy nowych przedsiębiorstw można wyjaśnić jedynie za pomocą wskaźników struktury przemysłu w regionie. W związku z tym konieczne jest w jego opinii uwzględnienie zarówno regionu, jak i branży wśród determinant wyjaśniających podjęcie decyzji o uruchomieniu i ewentualnie likwidacji przedsiębiorstwa.

Niezwykle ważna teoria, skupiająca się m.in. na opisie procesu powstawania przedsiębiorstw, została przedstawiona przez Jovanovica (1982). Według tej teorii proces ten jest w dużej mierze przypadkowy. Przypadkowość ta wynika z faktu, że – zakładając nową firmę – przedsiębiorca nie jest w stanie ocenić szans na jej przetrwanie, a nawet nie jest w stanie właściwie określić swoich kompetencji menedżerskich. Decyzja o pozostaniu na rynku lub o zakończeniu działalności wynika z konfrontacji niejasnych, mglistych oczekiwań przedsiębiorcy dotyczących działalności firmy z rzeczywistością.

Gerosky (1995) zwraca uwagę na to, że liczba nowych firm zależy od spodziewanych zysków w branży, ale zależność ta uwidacznia się dopiero od pewnego poziomu zysków. Do jego ciekawych spostrzeżeń należy zaliczyć to, że bariery rozpoczynania działalności gospodarczej są w praktyce barierami przetrwania na rynku i to właśnie nimi powinni zająć się badacze.

Teoria cyklu życia przedsiębiorstwa w swojej naturze nawiązuje do nauk biologicznych. Występuje tutaj analogia pomiędzy przedsiębiorstwami i organizmami żywymi, jeśli chodzi o ich fazy życia – narodziny, etap wzrostu, następnie stabilizacji, po której ewentualnie następuje starzenie się i śmierć. W nurcie nawiązującym do teorii biologicznych, a także teorii twórczej destrukcji Schumpetera (1999) mieści się praca Gurgula i inni (2014), w której za pomocą równania różniczkowego cząstkowego typu *von Foerstera* opisana jest dynamika powstawania i upadłości przedsiębiorstw niemieckich. Za pomocą tego modelu autorzy podjęli próbę prognozy liczby zlikwidowanych przedsiębiorstw i liczby przedsiębiorstw, które zostaną utworzone.

Literatura przedmiotu dotycząca cyklu życia przedsiębiorstw jest bogata. Levie, Lichtenstein (2010) twierdzą, że opublikowano już ponad 100 różnych modeli fazowych rozwoju przedsiębiorstwa. Liczba faz w tego typu modelach bywa różna, według nich waha się od trzech (np. Smith i inni, 1985) do jedenastu (np. Bruce, 1976).

Prusak (2011) analizował cykl życia przedsiębiorstwa pod kątem jego upadłości. Zaproponował schemat opisujący typowy przebieg cyklu życia, uwzględniający zmianę wartości przedsiębiorstwa, który składa się z czterech faz (Prusak, 2011, s. 52):

– założenie przedsiębiorstwa, zgromadzenie odpowiednich zasobów i rozpoczęcie działalności w wyniku sprzedaży produktów czy świadczenia usług,
– wzrost przedsiębiorstwa,
– stabilizacja w działalności przedsiębiorstwa,
– kryzys prowadzący do upadku przedsiębiorstwa lub kryzys będący etapem pośrednim w dalszym wzroście firmy.

Wyjątkowy charakter w modelu Prusaka ma faza pierwsza, która jako jedyna pojawia się tylko raz – zaraz po rozpoczęciu działalności gospodarczej. Z kolei czwarta faza cyklu (kryzys) prowadzić może do likwidacji przedsiębiorstwa lub do powrotu

do fazy drugiej (wzrost). W szczególności oznacza to, że zaproponowany przez niego model nie ma charakteru skończonego.

W modelu Greinera (1972) przyszłość przedsiębiorstwa jest zdeterminowana głównie jego historią, a nie czynnikami zewnętrznymi. W jego modelu rozwój firmy ma miejsce wskutek następujących naprzemiennie procesów ewolucji i rewolucji. Fazy ewolucji charakteryzuje stopniowy i zrównoważony wzrost, natomiast fazy rewolucji cechują się występowaniem wstrząsów i ogólnego zamieszania. Fazy rozwoju są mocno od siebie uzależnione i determinują się nawzajem. Podczas kryzysów w przedsiębiorstwie muszą zajść zmiany o charakterze organizacyjnym, które w myśl powiedzenia „co cię nie zabije, to cię wzmocni" przyczyniają się do jego rozwoju. Z drugiej strony brak odpowiedniego dostosowania się i wprowadzenia zmian w odpowiedzi na pojawiające się kryzysy skutkować może stagnacją i ewentualnie upadłością.

Według Greinera, Schein (1988) sprawność przedsiębiorstwa w początkowej fazie jego istnienia, wynika ze zdolności i kreatywności menedżera/właściciela. Dalsze ulepszenia wynikają już z problemów, na jakie organizacja napotyka w następnych etapach swojego istnienia i sposobów ich przezwyciężenia, co jest warunkiem jej przetrwania. Jeśli zarząd nie zauważy w porę tych problemów, to wcześniej czy później dochodzi do upadku tego przedsiębiorstwa. W warunkach niepewności, w których podejmuje się działania, w niezwykle szybko zmieniającym się otoczeniu, przy pojawiających się ciągle nowych problemach i zadaniach, cykl życia przedsiębiorstw w niezmienionej formie ulega w obecnej rzeczywistości gospodarczej znacznemu skróceniu. Oznacza to, że ma miejsce szybka eliminacja podmiotów, które trwają przy starych, niekonkurencyjnych już rozwiązaniach. Brak zmian wewnątrz przedsiębiorstwa prowadzi z upływem czasu do jego całkowitego upadku. Te zagadnienia są także przedmiotem rozważań Handy'ego (1996), który jest autorem tzw. „esowatej krzywej". Sednem jego tezy jest stwierdzenie, że przedsiębiorcy jeszcze w okresie prosperity powinni przygotowywać firmę na czas kryzysu, wywołując sztucznie objawy kryzysu i podejmując środki zaradcze. Choć firma na krótką metę wskutek tych sztucznych bodźców osłabia się, to jednak zmiany w odpowiedzi na wywołany kryzys na dłuższą metę wzmacniają i uodparniają tę firmę. Jest to efekt podobny do efektu szczepień ochronnych.

3. MODEL PRZYSPIESZONEJ PORAŻKI

Model przyspieszonej porażki (skrót AFT od ang. *accelerated failure time*) należy do grupy tzw. modeli parametrycznych. Istotą tych modeli jest założenie o znajomości postaci analitycznej funkcji gęstości prawdopodobieństwa rozkładu zmiennej losowej opisującej długość czasu trwania. Współczesne badania (Zając, 2013; Agarwal i inni, 2002; Cefis, Marsili, 2005) wskazują, że funkcja gęstości prawdopodobieństwa likwidacji przedsiębiorstw oraz związana z nią funkcja hazardu[3] nie są monotoniczne,

[3] Funkcja hazardu opisuje warunkową gęstość rozkładu długości czasu trwania przedsiębiorstwa w chwili *t*. Warunkiem jest założenie, że przedsiębiorstwo nadal istnieje w chwili *t*.

a ich kształt przypomina odwróconą literę „U". Ryzyko likwidacji firmy na początku wzrasta, by po osiągnięciu maksymalnego poziomu po kilku latach zmniejszyć się. W szczególności obserwuje się, że „tradycyjne" przedsiębiorstwa są najmniej narażone na ryzyko likwidacji[4]. Jovanovic (1982) argumentuje to faktem, że firmy potrzebują czasu, by nauczyć się w jaki sposób wykorzystywać szanse dające możliwość przetrwania i potencjalnego rozwoju. W związku z tym w prowadzonych badaniach empirycznych uwaga została zwrócona w kierunku U-kształtnych funkcji hazardu.

Głównym pomysłem, na którym oparte są modele przyspieszonej porażki, jest założenie, że zmienne egzogeniczne wpływają na długość czasu trwania. Oznacza to, że w wyniku oddziaływania zmiennych objaśniających funkcja przeżycia jest rozciągana (skracana) horyzontalnie. Własność rozciągania można zapisać jako zależność pomiędzy funkcjami przeżycia[5] dla dwóch różnych populacji $S_1(t)$ i $S_2(t)$. Model AFT zakłada bowiem, że istnieje stała ψ taka, że dla każdego t[6]:

$$S_1(t) = S_2(\psi t). \tag{1}$$

Klasycznym przykładem, pozwalającym zrozumieć opisaną tutaj zależność, jest przykład długości życia człowieka i psa. Przyjmuje się, że rok życia psa przekłada się na siedem lat życia człowieka, w związku z tym w tym przykładzie $\psi = 7$. Można zatem powiedzieć, że szansa na to by pies dożył 10 lat, jest taka jak szansa, by człowiek dożył 70 lat. Podobnie średnia długość życia psa jest siedem razy krótsza niż człowieka itp.

W sposób naturalny przyjmuje się, że wartość stałej ψ jest uzależniona od wartości zmiennych egzogenicznych. Zapisanie wzoru (1) z wykorzystaniem bazowej funkcji przeżycia daje zależność:

$$S(t|\mathbf{X}) = S_0(t\,\psi). \tag{2}$$

W praktyce do szacowania właściwych parametrów modelu AFT wykorzystuje się zależność opisaną równaniem:

$$\ln(t) = \boldsymbol{X}^T\boldsymbol{\beta} + \mu, \mu = c\varepsilon, \tag{3}$$

gdzie t oznacza czas, X^T jest macierzą zmiennych egzogenicznych, β to wektor parametrów modelu, natomiast μ jest składnikiem losowym. Składnik losowy ma formę zmiennej losowej ε pomnożonej przez skalującą ją stałą c. Rozkład zmiennej losowej ε jest powiązany z wybraną postacią funkcji gęstości prawdopodobieństwa likwidacji.

[4] Niektórzy wyciągają z tego wniosek, że podmioty powstałe w okresie rozkwitu gospodarczego mają największe szanse na przetrwanie na rynku (Gerosky i inni, 2010).

[5] Funkcja przeżycia (inaczej nazywana funkcją dożycia lub funkcją trwania) opisuje prawdopodobieństwo, że przedsiębiorstwo nie zostanie zlikwidowane do chwili t.

[6] Wzory (1)–(11) zostały zaczerpnięte z Landmesser (2013).

Umieszczenie obu stron równania (3) jako wykładników funkcji ekspotencjalnej, a następnie podzielenie równania stronami prowadzi do zależności:

$$t\,\mathrm{e}^{-\boldsymbol{X}^T\boldsymbol{\beta}} = \mathrm{e}^{\mu}. \tag{4}$$

Przyjmując oznaczenie $\psi = \mathrm{e}^{-\mathrm{X}^{\mathrm{T}}\beta}$, wzór (4) przybiera postać:

$$t\,\psi = \mathrm{e}^{\mu}. \tag{5}$$

Przy takich oznaczeniach współczynnik ψ standardowo nazywany jest parametrem przyśpieszającym (ang. *acceleration parameter*). Parametr ten jest stałą, o ile tylko zmienne egzogeniczne nie zmieniają się w czasie. Można powiedzieć, że „czas biegnie szybciej" dla obserwacji, dla których odpowiadające im ψ jest większe od jednostki (i analogicznie – wolniej, gdy $\psi < 1$).

Równanie (5) pozwala zrozumieć sposób parametryzacji przyjmowany przez praktyków. Niech τ oznacza czas trwania dla całej populacji, którego szybkość upływu jest traktowana jako norma i niech $\tau = t\psi$. Wtedy założenie, że ε jest zmienną losową prowadzi do wniosku, że podobnie jest z τ. Co więcej, jeżeli przykładowo ε ma rozkład normalny, to τ ma rozkład logarytmiczno-normalny[7].

Możliwe jest również zapisanie modelu AFT z wykorzystaniem funkcji hazardu. Wzór na warunkową funkcję hazardu w zależności od bazowej funkcji hazardu (czyli funkcji hazardu dla całej populacji), który będzie zgodny ze wzorem na funkcję przeżycia (2), przyjmuje wtedy postać:

$$h(t|\boldsymbol{X}) = \psi h_o(t\psi). \tag{6}$$

W celu dokonania właściwej interpretacji parametrów modelu, można chwilowo przyjąć, że dwie obserwacje o odpowiadających im wektorach X_i i X_j różnią się o jednostkę wartością zmiennych objaśniających na pojedynczej współrzędnej o indeksie p. Wówczas relację czasów trwania (ang. *time ratio*) dla tych obserwacji wyrazić można jako:

$$TR\left(t_i, t_j\right) = \frac{t_i}{t_j} = \frac{e^{\boldsymbol{X_i^T\beta}} e^{\varepsilon_i}}{e^{\boldsymbol{X_j^T\beta}} e^{\varepsilon_j}} \cong e^{(\boldsymbol{X}_i - \boldsymbol{X}_j)^T\boldsymbol{\beta}} = e^{\beta_p}. \tag{7}$$

Z powyższego wzoru wynika, że obliczona wartość $e^{\beta p}$ oznacza (przy założeniu *ceteris paribus*), ile razy zwiększyłby się średni czas trwania (długość życia) obserwowanego podmiotu, gdyby opisująca go zmienna objaśniająca x_p wzrosła o jednostkę.

[7] Równanie (3) jest zapisane w postaci liniowej, jednak ze względu na występowanie obserwacji cenzurowanych, nie jest możliwe szacowanie parametrów modelu klasyczną metodą najmniejszych kwadratów.

W badaniach empirycznych przeprowadzonych na potrzeby niniejszej pracy wykorzystane zostały trzy warianty modelu przyspieszonej porażki. W szczególności są to: model logarytmiczno-normalny, model log-logistyczny i uogólniony model gamma.

Zanim przejdziemy do opisu danych i estymacji omówionych modeli zwrócimy uwagę na zjawisko heterogeniczności podmiotów gospodarczych i jego uwzględnienie w modelach hazardu.

4. MODELE HAZARDU UWZGLĘDNIAJĄCE HETEROGENICZNOŚĆ

Tworząc probabilistyczny model hazardu, warto zwrócić uwagę na fakt, że poszczególne podmioty często nie spełniają założenia o homogeniczności, a funkcja hazardu dla populacji jest kombinacją hazardów indywidualnych jednostek (por. Gutierrez, 2002; Landmesser, 2013). Modele takie nazywa się modelami hazardu z heterogenicznością lub modelami z efektem *frailty*[8]. Różnice w wartościach funkcji hazardu dla poszczególnych obserwacji mogą mieć różnorakie źródła. Po pierwsze, mogą one wynikać z indywidualnych cech jednostek składających się na badaną zbiorowość. Różnice tego rodzaju w przypadku badania długości istnienia podmiotów gospodarki narodowej na rynku mogą być spowodowane jakością zarządzania firmą lub – w przypadku niewielkich firm – cechami charakteru właściciela. Drugą przyczyną występowania efektu *frailty* może być brak uwzględnienia wszystkich istotnych zmiennych egzogenicznych, które wpływają na funkcję hazardu. Trzecim powodem może być zbyt szybka likwidacja podmiotów znajdujących się w grupie wysokiego ryzyka lub istnienie podmiotów, o których można powiedzieć, że będą zawsze istnieć[9].

Brak próby uwzględnienia wpływu indywidualnych cech badanych podmiotów jest uważany za poważne niedociągnięcie, które może prowadzić do pozornych wniosków (Hougaard, 1984). Powszechnie przyjętym sposobem uwzględnienia efektu *frailty* w analizie przeżycia jest koncepcja modelu statystycznego, zawierającego dodatkowy efekt losowy przypisany każdej z obserwowanych jednostek, wpływający multiplikatywnie na długość okresu jej trwania (ang. *time-to-event*). Efekt ten można zatem traktować jako pewne uogólnienie dla parametrycznych modeli hazardu. Zmienność czasu trwania jest podzielona na część wyjaśnioną przez dostępne zmienne egzogeniczne i na pozostałą część. To właśnie ta druga część jest modelowana z wykorzystaniem losowego efektu *frailty*. Przewagą takiego podejścia przy znanym rozkładzie zmiennej losowej *frailty* jest otrzymanie warunkowej niezależności długości czasu trwania podmiotów.

Niech α_i oznacza efekt *frailty* związany z obserwacją o numerze i. Ponadto niech efekt ten oddziałuje multiplikatywnie na funkcję hazardu dla i-tej obserwacji:

---

[8] Ang. *frailty* – słabość, wątłość, niedociągnięcie, niedoskonałość, słabostka. W literaturze przedmiotu pojęcie nie jest zwykle tłumaczone na język polski, pojęcie po raz pierwszy wprowadzone w Vaupel i inni (1979).

[9] W tym sensie, że nie będzie możliwe zaobserwowanie ich likwidacji w rozsądnym/obserwowalnym przedziale czasowym.

$$h(t_i|\boldsymbol{X}_i, \alpha_i) = \alpha_i h(t_i|\boldsymbol{X}_i). \tag{8}$$

Przyjmuje się, że α_i jest realizacją zmiennej losowej α, która spełnia kilka założeń technicznych. Zmienna losowa jest niezależna zarówno od czasu t, jak i od wybranych zmiennych egzogenicznych. Dodatkowo jej realizacje są zawsze dodatnie i średnio nie wpływają na długość życia populacji ($E(\alpha) = 1$). Zakłada się ponadto, że jej wariancja jest skończona i oznaczona jako θ ($\sigma^2(\alpha) = \theta < \infty$). Można zatem zapisać, że:

$$h(t|\boldsymbol{X}, \alpha) = \alpha h(t|\boldsymbol{X}). \tag{9}$$

Warunkową funkcję przeżycia można teraz obliczyć, korzystając z (9) i (6):

$$S(t|\boldsymbol{X}, \alpha) = e^{-\int_0^t h(s|\boldsymbol{X},\alpha)ds} = e^{-\alpha\int_0^t h(s|\boldsymbol{X})ds} = e^{-\alpha H(t|\boldsymbol{X})} = e^{\alpha \ln(S(t|\boldsymbol{X}))} = [S(t|\boldsymbol{X})]^{\alpha}. \tag{10}$$

Ponieważ efekt *frailty* nie jest w praktyce obserwowalny, to funkcję przeżycia uzależnia się od parametrów rozkładu α, w szczególności od wariancji θ. Funkcję przeżycia, która zależy już nie od α, ale od wariancji α, otrzymuje się jako wartość oczekiwaną warunkowej funkcji przeżycia względem α, czyli wykorzystując (10) otrzymujemy:

$$S_\theta(t|\boldsymbol{X}) = \int_0^\infty S(t|\boldsymbol{X}, \alpha) g(\alpha) d\alpha = \int_0^\infty [S(t|\boldsymbol{X})]^{\alpha} g(\alpha) d\alpha, \tag{11}$$

gdzie przez $g(\alpha)$ rozumie się gęstość rozkładu prawdopodobieństwa zmiennej losowej α. W badaniach przeprowadzonych na potrzeby niniejszego artykułu przeanalizowane zostały warianty, w których efekt *frailty* był opisany rozkładem gamma i odwróconym rozkładem normalnym (skrót NIG od ang. *normal inverse Gaussian*)[10].

Estymację parametrów modelu przeprowadza się z wykorzystaniem metody największej wiarygodności. Aby sprawdzić, czy w modelu występuje efekt heterogeniczności, należy wyestymować parametry modelu uwzględniającego heterogeniczność, a następnie sprawdzić testem ilorazu wiarygodności Lagrange'a (lub testem Walda), czy estymator wariancji efektu *frailty* $\overline{\theta}$ jest statystycznie istotnie różny od zera. Zanim przejdziemy do estymacji przedstawimy wykorzystane w pracy dane.

5. DANE STATYSTYCZNE

Analiza empiryczna przeprowadzona w artykule bazuje na danych otrzymanych z Wojewódzkiego Urzędu Statystycznego w Krakowie. W szczególności na zbiór danych składają się informacje na temat wszystkich podmiotów gospodarki narodowej rejestrujących się w bazie REGON oraz podmioty wyrejestrowane z niej między 1 stycznia 2006 a 31 sierpnia 2014 w województwie małopolskim. W badanym okresie zarejestrowano łącznie 267 755 podmiotów gospodarki narodowej.

[10] Postać rozkładów przyjęta za Gutierrez (2002).

Specyfika posiadanego zbioru danych wymusza wykorzystanie tzw. obserwacji cenzurowanych prawostronnie. Przez obserwacje tego typu rozumie się w tym przypadku te dotyczące podmiotów gospodarki narodowej, które zostały zlikwidowane po 31 sierpnia 2014 lub nadal prowadzą działalność gospodarczą. Brak informacji na temat podmiotów gospodarki narodowej, które powstały przed analizowanym okresem i jednocześnie nie zakończyły działalności w czasie jego trwania, wymusił konkretny sposób uporządkowania zbioru danych. Dane posortowano w taki sposób, by wszystkie obserwowane podmioty rozpoczęły działalność w jednym, ustalonym momencie. Są one następnie obserwowane do momentu likwidacji lub do ostatniego dnia badania. W szczególności warto zauważyć, że prowadzi to do sytuacji, gdzie data cenzurowania jest najczęściej różna dla poszczególnych przedsiębiorstw.

Dla każdego z podmiotów możliwe jest wskazanie deklarowanej początkowej liczby pracowników (należy do jednej z pięciu grup: 0–9, 10–49, 50–249, 250–999, 1000 i więcej pracowników), sekcja PKD do której podmiot należy (PKD2004 i/lub PKD2007[11]) oraz miejsce prowadzenia działalności (powiat i rozróżnienie czy podmiot zarejestrowany jest w mieście czy na wsi).

Wykorzystanie tego typu danych na potrzeby modelu AFT wymaga wprowadzenia szeregu zmiennych dychotomicznych. W szczególności należy jednak pamiętać by poszczególne grupy otrzymane w ramach klastrów nie były zbyt małe. W związku z tym autorzy postanowili usunąć z badania przedsiębiorstwa deklarujące zatrudnienie ponad 250 pracowników (było ich zaledwie 16) oraz te należące do sekcji T[12] (jedynie 2) i U[13] (łącznie 6). Szczegółową listę pozostałych zmiennych zero-jedynkowych zamieszczono w tabeli 1[14].

Tabela 1.

Zmienne objaśniające wykorzystane w analizie parametrycznej

Nazwa zmiennej	Opis zmiennej, w nawiasie procent obserwacji, których dana wartość dotyczy	Nazwa zmiennej	Opis zmiennej, w nawiasie procent obserwacji, których dana wartość dotyczy
Mikro	Dla deklarowanego zatrudnienia 0–9 pracowników (98,7%)	p_1	powiat bocheński (2,5%)
Małe	Dla deklarowanego zatrudnienia 10–49 pracowników (1,2%)	p_2	powiat brzeski (2,0%)
Średnie	Dla deklarowanego zatrudnienia 50–249 pracowników (0,1%)	p_3	powiat chrzanowski (3,0%)

[11] W celu poprawy czytelności wyników autorzy zdecydowali się (tam gdzie to było konieczne) przekonwertować symbol PKD2004 do jego odpowiednika w PKD2007. W tym celu wykorzystano klucz powiązań publikowany przez GUS.

[12] Sekcja T oznacza „gospodarstwa domowe zatrudniające pracowników; gospodarstwa domowe produkujące wyroby i świadczące usługi na własne potrzeby".

[13] Sekcja U oznacza „organizacje i zespoły eksterytorialne".

[14] W odczuciu autorów wykorzystanie tak dużej liczby zmiennych dychotomicznych jest uzasadnione wielkością posiadanego zbioru danych.

Tabela 1. (cd.)

Nazwa zmiennej	Opis zmiennej, w nawiasie procent obserwacji, których dana wartość dotyczy	Nazwa zmiennej	Opis zmiennej, w nawiasie procent obserwacji, których dana wartość dotyczy
A	Rolnictwo, leśnictwo, łowiectwo i rybactwo (0,7%)	p_4	powiat dąbrowski (1,0%)
B	Górnictwo i wydobywanie (0,1%)	p_5	powiat gorlicki (2,8%)
C	Przetwórstwo przemysłowe (7,8%)	p_6	powiat krakowski (7,5%)
D	Wytwarzanie i zaopatrywanie w energię elektryczną, gaz, parę wodną, gorącą wodę i powietrze do układów klimatyzacyjnych (0,2%)	p_7	powiat limanowski (3,4%)
E	Dostawa wody; gospodarowanie ściekami i odpadami oraz działalność związana z rekultywacją (0,3%)	p_8	powiat miechowski (1,0%)
F	Budownictwo (17,8%)	p_9	powiat myślenicki (3,3%)
G	Handel hurtowy i detaliczny; naprawa pojazdów samochodowych, włączając motocykle (24,5%)	p_{10}	powiat nowosądecki (5,5%)
H	Transport i gospodarka magazynowa (5,4%)	p_{11}	powiat nowotarski (5,1%)
I	Działalność związana z zakwaterowaniem i usługami gastronomicznymi (4,5%)	p_{12}	powiat olkuski (2,8%)
J	Informacja i komunikacja (3,8%)	p_{13}	powiat oświęcimski (3,6%)
K	Działalność finansowa i ubezpieczeniowa (3,8%)	p_{14}	powiat proszowicki (0,8%)
L	Działalność związana z obsługą rynku nieruchomości (2,2%)	p_{15}	powiat suski (1,9%)
M	Działalność profesjonalna, naukowa i techniczna (9,0%)	p_{16}	powiat tarnowski (3,9%)
N	Działalność w zakresie usług administrowania i działalność wspierająca (4,2%)	p_{17}	powiat tatrzański (2,7%)
O	Administracja publiczna i obrona narodowa; obowiązkowe zabezpieczenia społeczne (0,04%)	p_{18}	powiat wadowicki (3,9%)
P	Edukacja (3,5%)	p_{19}	powiat wielicki (3,6%)
Q	Opieka zdrowotna i pomoc społeczna (4,4%)	p_{61}	miasto Kraków (33,1%)
R	Działalność związana z kulturą, rozrywką i rekreacją (1,8%)	p_{62}	miasto Nowy Sącz (3,1%)
S	Pozostała działalność usługowa (6,1%)	p_{63}	miasto Tarnów (3,4%)
Miasto	Dla podmiotów zarejestrowanych w miastach (61,0%)		

Źródło: opracowanie własne, na podstawie danych uzyskanych z Urzędu Statystycznego w Krakowie.

Przed przystąpieniem do właściwej analizy parametrycznej należy uświadomić sobie występowanie ścisłej współliniowości wśród zdefiniowanych w tabeli 1 zmiennych. Współliniowość ta bierze się z faktu, że każda z obserwacji musi przynależeć do jednaj z klas wielkości, mieć przypisany kod odpowiadającej jej sekcji PKD oraz należeć do jednego z powiatów województwa. Zapisanie tego spostrzeżenia za pomocą aparatu matematycznego prowadzi do zależności:

$$\begin{cases} Mikro + Małe + Średnie = 1, \\ A + \cdots + S = 1, \\ p_1 + \cdots + p_{63} = 1. \end{cases} \tag{12}$$

Sytuacja ta wymusza wskazanie jednego z klastrów jako punktu odniesienia dla pozostałych. Jako kryterium wybrano do tego celu wielkość poszczególnych podgrup. Punktem odniesienia są więc mikroprzedsiębiorstwa, których główna działalność gospodarcza wskazuje na ich przynależność do sekcji F, zarejestrowane w Krakowie. Korzystając z zaprezentowanych danych wykonano estymację omówionych wcześniej modeli. Jej wyniki są zamieszczone w następnym rozdziale.

6. WYNIKI ESTYMACJI MODELI PRZYSPIESZONEJ PORAŻKI AFT

W niniejszym artykule rozważane są trzy zasadnicze typy modeli AFT – model logarytmiczno-normalny, model log-logistyczny i uogólniony model gamma. Aby wybrać wersję modelu, która będzie najlepiej pasować do danych, wykorzystano kryterium informacyjne Akaike (AIC) i kryterium bayesowskie Schwarza (BIC).

Rezultaty estymacji zostały zamieszczone w tabeli 2. Można zauważy, że zarówno kryterium AIC jak i kryterium BIC wskazują na model log-logistyczny jako na ten, który jest najlepiej odpowiada danym empirycznym. W związku z tym dla modelu log-logistycznego uwzględnione zostało występowanie efektu *frailty*. Efekt ten został przebadany w dwóch wersjach. W pierwszej efekt *frailty* ma rozkład gamma, a w drugiej odwrócony rozkład normalny NIG. Wartości kryteriów informacyjnych w obu przypadkach są do siebie mocno zbliżone, ale oba wskazują na rozkład gamma jako lepiej pasujący do danych.

Analizując wyniki, warto zwrócić uwagę na podobne wartości parametrów otrzymane dla poszczególnych zmiennych we wszystkich wyestymowanych wersjach modelu AFT. Nie licząc dwóch wyjątków, parametry są statystycznie istotnie różne od zera, brak istotności dotyczy zmiennych p_{17} (powiat tatrzański) i p_{19} (powiat wielicki). Dla tych dwóch powiatów można przyjąć, że średni czas przeżycia podmiotów gospodarki narodowej jest niemal taki sam jak w Krakowie.

Tabela 2.

Wyniki estymacji modeli przyspieszonej porażki AFT

Zmienna	Model logistyczno-normalny	Uogólniony model gamma	Model Log-logistyczny		
			Bez efektu *frailty*	NIG	Gamma
Małe	-0,929 ***	-0,963 ***	-0,963 ***	-0,961 ***	-0,960 ***
Średnie	-1,848 ***	-1,804 ***	-1,768 ***	-1,761 ***	-1,760 ***
A	-0,311 ***	-0,414 ***	-0,396 ***	-0,385 ***	-0,383 ***
B	-0,931 ***	-0,937 ***	-0,918 ***	-0,912 ***	-0,911 ***
C	-0,217 ***	-0,199 ***	-0,192 ***	-0,191 ***	-0,190 ***
D	-0,433 ***	-0,484 ***	-0,452 ***	-0,439 ***	-0,437 ***
E	-0,283 ***	-0,356 ***	-0,335 ***	-0,325 ***	-0,326 ***
G	-0,177 ***	-0,130 ***	-0,151 ***	-0,157 ***	-0,158 ***
H	-0,112 ***	-0,048 ***	-0,076 ***	-0,086 ***	-0,088 ***
I	-0,224 ***	-0,119 ***	-0,154 ***	-0,168 ***	-0,170 ***
J	-0,475 ***	-0,434 ***	-0,420 ***	-0,418 ***	-0,417 ***
K	-0,464 ***	-0,361 ***	-0,401 ***	-0,413 ***	-0,414 ***
L	-0,478 ***	-0,491 ***	-0,472 ***	-0,466 ***	-0,465 ***
M	-0,646 ***	-0,619 ***	-0,605 ***	-0,602 ***	-0,601 ***
N	-0,284 ***	-0,301 ***	-0,272 ***	-0,262 ***	-0,261 ***
O	-0,957 ***	-1,063 ***	-1,043 ***	-1,032 ***	-1,031 ***
P	-0,485 ***	-0,498 ***	-0,481 ***	-0,475 ***	-0,474 ***
Q	-0,749 ***	-0,793 ***	-0,778 ***	-0,772 ***	-0,771 ***
R	-0,346 ***	-0,362 ***	-0,338 ***	-0,330 ***	-0,329 ***
S	-0,143 ***	-0,212 ***	-0,178 ***	-0,165 ***	-0,163 ***
p_1	-0,046 ***	-0,056 ***	-0,057 ***	-0,057 ***	-0,057 ***
p_2	-0,165 ***	-0,161 ***	-0,168 ***	-0,169 ***	-0,169 ***
p_3	-0,150 ***	-0,157 ***	-0,162 ***	-0,162 ***	-0,162 ***
p_4	-0,287 ***	-0,295 ***	-0,312 ***	-0,314 ***	-0,314 ***
p_5	-0,237 ***	-0,240 ***	-0,242 ***	-0,242 ***	-0,241 ***
p_6	-0,054 ***	-0,042 ***	-0,046 ***	-0,047 ***	-0,047 ***
p_7	-0,296 ***	-0,303 ***	-0,310 ***	-0,310 ***	-0,310 ***

Zmienna	Model logistyczno-normalny	Uogólniony model gamma	Model Log-logistyczny		
			Bez efektu *frailty*	NIG	Gamma
p_8	-0,144 ***	-0,147 ***	-0,156 ***	-0,157 ***	-0,157 ***
p_9	-0,054 ***	-0,074 ***	-0,078 ***	-0,078 ***	-0,078 ***
p_{10}	-0,247 ***	-0,239 ***	-0,245 ***	-0,246 ***	-0,246 ***
p_{11}	-0,471 ***	-0,412 ***	-0,449 ***	-0,460 ***	-0,461 ***
p_{12}	-0,131 ***	-0,152 ***	-0,151 ***	-0,150 ***	-0,149 ***
p_{13}	-0,074 ***	-0,080 ***	-0,079 ***	-0,078 ***	-0,078 ***
p_{14}	-0,085 ***	-0,085 ***	-0,087 ***	-0,088 ***	-0,088 ***
p_{15}	-0,130 ***	-0,127 ***	-0,133 ***	-0,135 ***	-0,135 ***
p_{16}	-0,170 ***	-0,170 ***	-0,176 ***	-0,178 ***	-0,178 ***
p_{17}	-0,007 ***	-0,008 ***	-0,001 ***	-0,003 ***	-0,003 ***
p_{18}	-0,157 ***	-0,146 ***	-0,159 ***	-0,163 ***	-0,163 ***
p_{19}	-0,031 ***	-0,019 ***	-0,022 ***	-0,023 ***	-0,023 ***
p_{62}	-0,291 ***	-0,286 ***	-0,287 ***	-0,286 ***	-0,286 ***
p_{63}	-0,282 ***	-0,270 ***	-0,276 ***	-0,278 ***	-0,278 ***
Miasto	-0,081 ***	-0,079 ***	-0,085 ***	-0,086 ***	-0,087 ***
const	-7,861 ***	-8,006 ***	-7,744 ***	-7,707 ***	-7,703 ***
	σ = 1,617	σ = 1,172 κ = 0,675	γ = 0,844	γ = 0,826 θ = 0,124	γ = 0,824 θ = 0,133
Loglik *AIC* *BIC*	-244686,1 -489460,2 -489921,3	-242277,2 -484889,0 -485360,6	-242008,8 -484105,7 -484566,8	-241989,7 -484069,4 -484541,1	-241987,0 -484064,3 -484535,9
				LR = 38,26 ***	*LR* = 43,40 ***

Oznaczenia: *** poziom istotności 1%; ** poziom istotności 5%; * poziom istotności 10%.

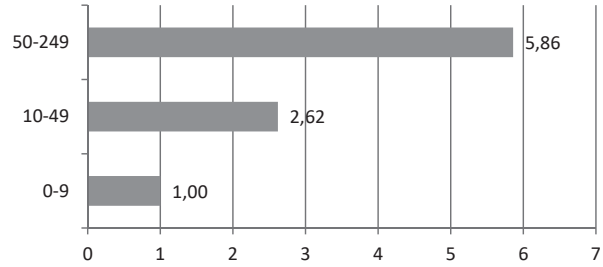
Źródło: opracowanie własne.

Jako przykład właściwej interpretacji otrzymanej wartości parametrów posłuży zmienna *Małe*, która przyjmuje wartość 1 dla podmiotów deklarujących zatrudnienie od 10 do 49 pracowników i 0 w przeciwnym przypadku. Parametr otrzymany dla tej zmiennej jest dodatni, co oznacza, że czas trwania dla przedsiębiorstw należących do tej grupy jest średnio dłuższy niż dla mikroprzedsiębiorstw (0–9 pracowników) traktowanych w tym przypadku jako punkt odniesienia. Wartość przyjmowana przez parametr wynosi 0,963, w związku z tym firmy małe istnieją średnio 2,6 ($e^{0,963} = 2,620$)

razy dłużej niż mikroprzedsiębiorstwa. Ujemne wartości wyestymowanego parametru interpretuje się w ten sposób, że czas trwania dla podmiotów należących do danej podgrupy jest krótszy niż dla tych będących punktem odniesienia. Jeżeli wartość parametru jest równa zero, to czas trwania jest równy temu bazowemu.

Punktem odniesienia w analizie są podmioty mające między 0 a 9 pracowników, zarejestrowane w Krakowie w sekcji F. Oczekiwany czas trwania w rejestrze REGON dla podmiotów spełniających te warunki wynosi $e^{7,744 + 0,085} = 2512,416$ dni (w potędze znajduje się stała i premia związana z faktem zarejestrowania w mieście).

Ogólne wyniki zobrazowane zostaną graficznie. Jako pierwszy przeanalizowany zostanie podział ze względu na wielkość mierzoną deklarowaną liczbą pracowników (rysunek 1).

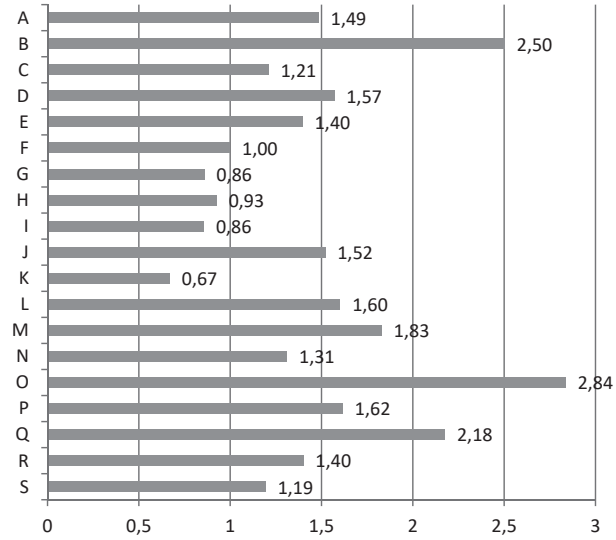


Rysunek 1. Czas trwania względem wielkości bazowej dla podmiotów zarejestrowanych w bazie REGON – podział ze względu na liczbę pracowników

Źródło: opracowanie własne.

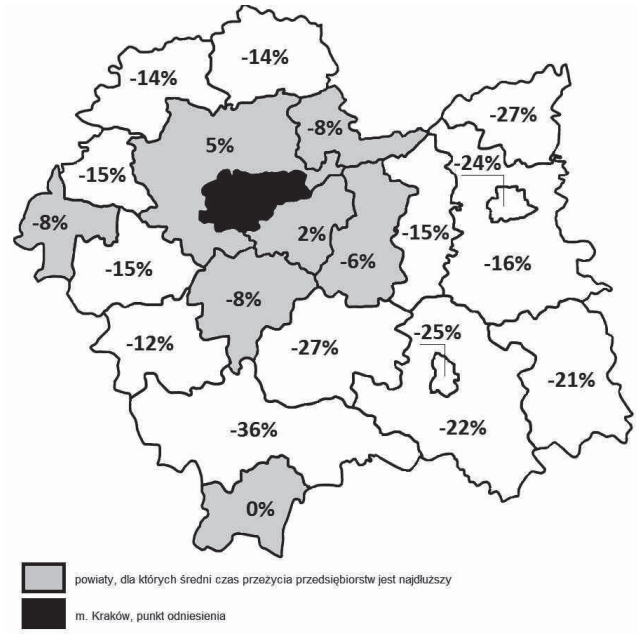
Wyraźnie zauważalne jest wydłużanie się czasu przeżycia, które następuje wraz ze wzrostem liczby pracowników. Czas trwania firm małych jest ponad dwukrotnie dłuższy niż mikroprzedsiębiorstw, a z kolei czas trwania firm średnich jest ponad dwukrotnie większy niż firm małych.

Podział ze względu na sekcję PKD jest również wyraźny (rysunek 2). Najkrótszy czas trwania przypada podmiotom gospodarki narodowej zarejestrowanym w sekcji K (działalność finansowa i ubezpieczeniowa). Czas trwania dla tej grupy jest o jedną trzecią krótszy od czasu trwania w największej sekcji F. Następnymi w kolejności są sekcje I, G oraz H. Najdłuższy czas trwania przypada (zgodnie z przewidywaniami) sekcji O (administracja), zaraz za nią znajduje się sekcja B (górnictwo). Czas trwania dla podmiotu zarejestrowanego w sekcji B jest 2,5 razy dłuższy niż dla firm z sekcji F. Czas trwania podmiotów gospodarki narodowej zarejestrowanych w sekcjach A, D, E, J, L, P i R jest porównywalny i o połowę dłuższy od czasu trwania podmiotów z sekcji F.



Rysunek 2. Czas trwania względem wielkości bazowej dla podmiotów zarejestrowanych w bazie REGON – podział ze względu na sekcję PKD2007

Źródło: opracowanie własne.



Rysunek 3. Czas trwania względem wielkości bazowej dla podmiotów zarejestrowanych w bazie REGON – podział ze względu powiat, w którym podmiot został zarejestrowany

Źródło: opracowanie własne.

Interpretując parametry zmiennych związanych z podziałem ze względu na powiat, w którym podmiot został zarejestrowany, zauważyć można wyraźne zróżnicowanie związane z odległością od Krakowa (rysunek 3). Powiaty sąsiadujące z miastem wojewódzkim cechują się najdłuższą przeżywalnością, nawet o kilka procent większą niż w samym Krakowie. Wraz ze wzrostem odległości od Krakowa czas funkcjonowania podmiotów skraca się. Wyjątkiem jest powiat tatrzański, w którym długość trwania podmiotów pomimo względnie dużej odległości od Krakowa jest taka sama jak w stolicy Małopolski. Północno-zachodnia część województwa wyróżnia się dłuższym czasem trwania podmiotów gospodarczych niż południowo-wschodnia. Najkrócej istnieją podmioty zarejestrowane w powiecie nowotarskim, gdzie czas trwania jest o jedną trzecią krótszy niż w Krakowie.

Długość okresu istnienia podmiotów miejskich i wiejskich okazuje się również istotnie różnić między sobą. Oczekiwany okres istnienia podmiotów zarejestrowanych w miastach jest o niemal 9% dłuższy od czasu istnienia ich odpowiedników zarejestrowanych na wsi. Należy o tym wpływie pamiętać interpretując długość trwania podmiotów zakładanych w powiatach miejskich (krakowskim, tarnowskim i nowosądeckim).

Można przyjąć, że wybrany model log-logistyczny cechuje się nieobserwowalną heterogenicznością (tabela 2). Celem uchwycenia nieobserwowalnej heterogeniczności zastosowano modele uwzględniające efekt *frailty*, czyli efekt oddziałujący multiplikatywnie na czas trwania, przypisany indywidualnie każdej z badanych jednostek. Wykorzystanie w tym celu rozkładu NIG oraz rozkładu gamma dało istotne statystycznie rezultaty. W obu przypadkach test ilorazu wiarygodności odrzuca hipotezę zerową o braku istnienia efektu *frailty*. Modelem wskazywanym jako lepszy zarówno przez kryterium AIC jak i kryterium BIC jest model, w którym efekt *frailty* pochodzi z rozkładu gamma. W szczególności estymator wariancji $Var(\hat{\theta}) = 0{,}133$, czyli estymator odchylenia standardowego wynosi 0,365. Oznacza to, że indywidualne cechy przedsiębiorstwa, które odzwierciedlają jakość zarządzania firmą, wpływają na wydłużenie lub skrócenie czasu funkcjonowania na rynku średnio o 36%.

Występowanie nieobserwowalnej heterogeniczności ma również wpływ na interpretację oszacowanych parametrów modelu. Interpretując parametry modelu, należy pamiętać, że wydłużenie/skrócenie czasu trwania, dotyczy podmiotów o tej samej wartości efektu *frailty*.

7. WNIOSKI

Wykonana analiza pozwala zobaczyć, w jaki sposób wielkość przedsiębiorstwa mierzona liczbą pracowników, rodzaj prowadzonej działalności gospodarczej oraz miejsce zarejestrowania wpływa na długość trwania firmy. Badanie wykonano na grupie 267 tys. podmiotów gospodarki narodowej zakładanych w województwie małopolskim w latach 2006–2014. Zgodnie z najlepszą wiedzą autorów badania tego typu prowadzone z wykorzystaniem modeli przyspieszonej porażki zostały wykonane po raz pierwszy właśnie w tej pracy.

Średnia długość istnienia przedsiębiorstw jest w sposób statystycznie istotny uzależniona od liczby zatrudnianych pracowników. Przedsiębiorstwa średnie istnieją statystycznie ponad dwukrotnie dłużej od małych, natomiast małe znowu ponad dwukrotnie dłużej od mikro.

Rodzaj prowadzonej działalności gospodarczej również ma znaczący wpływ na średnią oczekiwaną długość czasu trwania na rynku. Najkrócej działają firmy zakwalifikowane do sekcji K, czyli firmy zajmujące się działalnością finansową i ubezpieczeniową, najdłużej zaś te, których dominująca działalność gospodarcza kwalifikuje do sekcji B, czyli górnictwo i wydobywanie, oraz O, czyli administracja publiczna i obrona narodowa; obowiązkowe zabezpieczenia społeczne. Podmioty zakwalifikowane do sekcji B i O istnieją średnio czterokrotnie dłużej od tych z sekcji K.

Analizując wpływ lokalizacji działalności gospodarczej na długość trwania na rynku zauważalna jest wyraźnie zależność, że im większa odległość od Krakowa tym krótszy okres działania firmy. Najkrócej istnieją firmy zakładane w powiecie nowotarskim, a najdłużej w powiecie krakowskim. Zalety turystyczne powiatu tatrzańskiego przekładają się istotnie na wydłużenie średniej długości istnienia firm tam zakładanych. Firmy zakładane w miastach istnieją średnio o 9% dłużej niż te zakładane na wsi.

Badania empiryczne oparte na modelach przyspieszonej porażki jednoznacznie wskazują na występowanie zjawiska heterogeniczności funkcji hazardu, a więc potwierdzają istotność indywidualnych cech przedsiębiorstwa (wynikających np. z jakości zarządzania firmą), które (według badań) odpowiadają za wydłużenie/skrócenie spodziewanego czasu trwania na rynku średnio o ok. 36%.

LITERATURA

Agarwal R., Sarkar M. B., Echambadi R., (2002), The Conditioning Effect of Time on Firm Survival: An Industry Life Cycle Approach, *Academy of Management Journal*, 45 (5), 971–994.

Allison P. D., (1995), *Survival Analysis Using Sas: A Practical Guide*, Sas Institute Inc., Cary.

Armington C., Acs Z., (2002), The Determinants of Regional Variation in New Firm Formation, *Regional Studies*, 36 (1), 33–45.

Bruce R., (1976), *The Entrepreneurs: Strategies, Motivations, Successes, and Failures*, U.K. Libertarian Books, Bedford.

Cefis E., Marsili O., (2005), A Matter of Life and Death: Innovation and Firm Survival, *Industrial and Corporate Change*, 14 (6), 1–26.

Fritsch M., (1997), New Firms and Regional Employment Change, *Small Business Economics*, 2, 437–448.

Gerosky P. A., (1995), What do we Know About Entry, *International Journal of Industrial Organization*, 13 (4), 413–614.

Gerosky P. A., Mata J., Portugal P., (2010), Founding Conditions and the Survival of New Firms, *Strategic Management Journal*, 31 (5), 510–529.

Greiner L. E., (1972), Evolution and Revolution as Organizations Grow, *Harvard Business Review*, 50 (4), 37–46.

Greiner L. E., Schein V. E., (1988), *Power and Organization Development: Mobilizing Power to Implement Change*, Addison-Wesley Publishing Company, New York.

Gurgul H., Zając P., Matschke X., Matschke M. J., (2014), A Dynamic Model of Birth and Death of Enterprises in Germany, *Betriebswirtschaftliche Forschung und Praxis*, 66 (1), 86–103.

Gutierrez R. G., (2002), Parametric Frailty and Shared Frailty Survival Models, *The Stata Journal*, 2 (1), 22–44.
Handy C., (1996), *Wiek paradoksu*, Dom Wydawniczy ABC, Warszawa.
Hougaard P., (1984), Life Table Methods for Heterogeneous Populations: Distributions Describing the Heterogeneity, *Biometrika*, 71 (1), 75–83.
Jovanovic B., (1982), Selection and Evolution of Industry, *Econometrica*, 50 (3), 649–670.
Knight F. H., (1921), Risk, Uncertainty, and Profit, *Hart, Schaffner, and Marx Prize Essays*, nr. 31, Boston i New York: Houghton Mifflin.
Landmesser J., (2013), *Wykorzystanie metod analizy czasu trwania do badania aktywności ekonomicznej ludności w Polsce*, Wydawnictwo SGGW, Warszawa.
Levie J., Lichtenstein B. B., (2010), A Terminal Assessment of Stages Theory: Introducing a Dynamic States Approach to Entrepreneurship, *Entrepreneurship Theory and Practice*, 34 (2), 317–350.
Markowicz I., (2012), *Statystyczna analiza żywotności firm*, Wydawnictwo Naukowe Uniwersytetu Szczecińskiego, Szczecin.
Nehrebecka N., Dzik A. M. (2013), Business Demography in Poland: Microeconomic and Macroeconomic Determinants of Firm Survival, *Working Papers 2013-08*, Faculty of Economic Sciences, University of Warsaw.
Pociecha J., (2012), Model logitowy jako narzędzie prognozowania bankructwa. Jego zalety i ograniczenia, w: *Spotkania z królową nauk*, Wydawnictwo Uniwersytetu Ekonomicznego, Kraków.
Prusak B., (2011), *Ekonomiczna analiza upadłości przedsiębiorstw – ujęcie międzynarodowe*, CeDeWu, Warszawa.
Ptak-Chmielewska A., (2010), Analiza przeżycia przedsiębiorstw w Polsce na przykładzie wybranego województwa, w: Dittmann P., (red.), *Prognozowanie w zarządzaniu firmą*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław.
Schumpeter J. A., (1999), *Teoria rozwoju gospodarczego*, PWE, Warszawa.
Smith K., Mitchell T., Summer C., (1985), Top Level Management Priorities in Different Stages of the Organization Life Cycle, *Academy of Management Journal*, 28 (4), 799–820.
van Praag C. M., (1996), *Determinants of Successful Entrepreneuership*, Thesis Publisher, Amsterdam.
Vaupel J. W., Manton K. G., Stallard E., (1979), The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality, *Demography*, 16 (3), 439–454.
Zając P., (2013), The New Approach to Estimation of the Hazard Function in Business Demography on Example of Data from New Zealand, *Managerial Economics*, 13, 99–110.

ZASTOSOWANIE MODELU PRZYSPIESZONEJ PORAŻKI DO ANALIZY PRZEŻYCIA MAŁOPOLSKICH PRZEDSIĘBIORSTW

Streszczenie

Celem artykułu było przeprowadzenie analizy przeżycia przedsiębiorstw rejestrowanych na terenie województwa małopolskiego w latach 2006–2014. Wykorzystany w tym celu zbiór danych zawierał informacje na temat ponad 267 tys. podmiotów gospodarki narodowej i posłużył zbadaniu wpływu wybranych czynników na długość czasu funkcjonowania na rynku. Zgodnie z wiedzą autorów są to jedne z pierwszych badań tego typu na obszarze Europy Środkowo-Wschodniej w których wykorzystano modele przyspieszonej porażki. Istotnymi czynnikami mającymi wpływ na długość okresu istnienia firmy jest jego wielkość, rodzaj działalności gospodarczej oraz miejsce zarejestrowania. Przedsiębiorstwa średnie istnieją statystycznie ponad dwukrotnie dłużej od małych, natomiast małe ponad dwukrotnie dłużej od mikro. Najkrócej działają firmy zajmujące się działalnością finansową i ubezpieczeniową, zaś najdłużej (czterokrotnie dłużej) firmy z obszarów górnictwo i wydobywanie surowców oraz administracja publiczna

i obrona narodowa, a także obowiązkowe zabezpieczenia społeczne. Badania wykazały, że im większa odległość od Krakowa tym krótszy okres działania firmy. Firmy zakładane w miastach działają średnio o 9% dłużej niż te zakładane na wsi. Badania za pomocą modeli przyspieszonej porażki jednoznacznie wskazują na występowanie zjawiska heterogeniczności funkcji hazardu.

Słowa kluczowe: przedsiębiorstwa, Małopolska, model przyspieszonej porażki, funkcja hazardu, analiza przeżycia

THE APPLICATION OF AN ACCELERATED FAILURE TIME MODEL TO THE SURVIVAL ANALYSIS OF ENTERPRISES FOUNDED IN LESSER POLAND VOIVODSHIP

Abstract

The goal of this research was the survival analysis of enterprises founded in Lesser Poland Voivodship between years 2006–2014. The sample comprising 267 thousands firms was used to examine whether certain factors such as the size of the company, area of its activity and registration place had statistically significant impact on their duration time. All of those proved to be substantial in the course of our study. In general companies belonging to the sector of finance and insurance survived the shortest, whereas the longest duration (even four times longer than in finance and insurance) was among companies from mining sector, public administration, national defense and social insurance. Moreover, the research has proved that while moving further from Cracow, the survival of companies was likely to shorten. The only exception was Tatra Country, where duration was relatively long despite the distance factor. Additionally companies in urban areas were active approximately 9% longer than firms founded in rural ones. To the best of our knowledge this is the first application of an AFT (*accelerated failure time*) model to survival analysis of companies based on the data from Central European country. In this case its application proves existence of the heterogeneity of the hazard function.

Keywords: enterprises, Lesser Poland Voivodeship, accelerated failure time model, hazard function, survival analysis

DOI: 10.5604/01.3001.0014.1223

PRZEGLĄD STATYSTYCZNY
R. LXIII – ZESZYT 4 – 2016

JOANNA KISIELIŃSKA[1]

ROZKŁADY WYBRANYCH BOOTSTRAPOWYCH ESTYMATORÓW MEDIANY ORAZ ZASTOSOWANIE DOKŁADNEJ METODY PERCENTYLI DO JEJ PRZEDZIAŁOWEGO SZACOWANIA

1. WPROWADZENIE

Do szacowania parametrów rozkładu zmiennej losowej w przypadku, gdy rozkład ten nie jest znany, można zastosować metodę bootstrapową. Metoda, zaproponowana przez Efrona (1979), polega na wtórnym próbkowaniu próby losowej (nazywanej dalej pierwotną) i przybliżaniu rozkładu estymatora parametru rozkładem określonym dla prób wtórnych (bootstrapowym rozkładem estymatora). Liczba możliwych do wylosowania (ze zwracaniem) ze skończonej próby pierwotnej, prób wtórnych jest skończona i równa liczbie wariacji z powtórzeniami. Jeśli próba pierwotna liczy n elementów, prób wtórnych jest n^n. Wiele z nich różni się jedynie kolejnością elementów – prób wtórnych zawierających różne elementy jest $\binom{2n-1}{n}$ (Fisher, Hall, 1991).

Klasyczną metodę bootstrapową można w uproszczeniu interpretować jako losowanie ze zwracaniem N-elementowej próby ze zbioru wszystkich prób wtórnych, liczącego $B = n^n$ elementów. W przypadku małych prób pierwotnych może się wiec zdarzyć, że liczba losowanych prób wtórnych N będzie większa od B i liczbę losowanych prób wtórnych trzeba ograniczyć. Ze względu na skończoną liczbę prób wtórnych możliwe jest ponadto wielokrotne wylosowanie tej samej próby, z kolei inna może nie być wylosowana ani razu. Określając rozkład estymatora bootstrapowego warto więc wykorzystywać wszystkie możliwe próby wtórne. Metodę bazującą na wszystkich próbach wtórnych nazywa się dokładną metodą bootsrapową (*exact bootstrap method*). Metoda przedstawiona została między innymi przez Fishera, Halla (1991), Kisielińską (2013). Na potrzebę stosowania metody dokładnej zwracają uwagę Hutson, Ernst (2000). Zauważają bowiem, że metoda ta pozwala wyeliminować błędy wynikające z losowania prób wtórnych.

Należy zwrócić uwagę, że dokładna metoda bootstrapowa nazywa się dokładną jedynie dlatego, że wykorzystuje wszystkie możliwe próby wtórne. Kwestia dokładności szacunku parametru badanej populacji względem jego wartości faktycznej dotyczy próby pierwotnej i zasad jej pobierania.

[1] Szkoła Główna Gospodarstwa Wiejskiego, Wydział Nauk Ekonomicznych, Katedra Ekonomiki Rolnictwa i Międzynarodowych Stosunków Gospodarczych, ul. Nowoursynowska 166, 02-787 Warszawa, Polska, e-mail: joanna_kisielinska@sggw.pl.

Parametrem najczęściej wykorzystywanym do charakterystyki poziomu zjawiska jest wartość oczekiwana. Jednak w przypadku, gdy rozkład cechuje asymetria, warto zastosować medianę, której wybrane metody punktowego i przedziałowego szacowania są przedmiotem niniejszego artykułu. Evans i inni (2006) przedstawili wzory pozwalające wyznaczyć rozkłady bootstrapowych estymatorów statystyk pozycyjnych o dowolnej randze. Mogą być one zastosowane dla określenia rozkładu estymatora w postaci mediany z próby o nieparzystej liczbie obserwacji (estymator standardowy). Nie można ich jednak użyć, do wyznaczenia rozkładu estymatora w postaci mediany z próby oraz estymatora standardowego dla próby o parzystej liczbie obserwacji. W pracy zaproponowane zostaną formuły określające rozkład estymatora standardowego (dla próby o parzystej liczbie elementów). Dla estymatora w postaci mediany z próby o parzystej liczbie elementów pokazane zostanie, że nie jest możliwe podanie wzorów ogólnych (obowiązujących dla każdej próby) określających ten rozkład. W pracy pokazane zostanie, jak mimo tego rozkład estymatora wyznaczyć.

W artykule zwrócono uwagę na istotne własności rozkładu standardowego estymatora mediany (zarówno dla próby o nieparzystej jak i parzystej liczbie elementów), które pozwalają stworzyć gotowe tablice oraz z góry określić, które elementy próby pierwotnej są granicami przedziałów ufności. W przypadku estymatora w postaci mediany z próby o parzystej liczbie elementów, nie jest to możliwe i dla każdej próby należy przedziały wyznaczać indywidualne.

Podkreślić należy, że przedziały ufności dla trzech wziętych pod uwagę estymatorów wyznaczane były metodą percentyli zastosowaną dla wszystkich możliwych prób wtórnych (wykorzystano dokładną metodę bootstrapu). Metodę taką można nazwać dokładną metodą percentyli przez analogię do dokładnej metody bootstrapowej. W pracy pokazane zostanie, że w przypadku standardowego estymatora mediany określanie granic przedziałów ufności dokładną metodą percentyli jest znacznie prostsze niż zastosowanie metody klasycznej (z losowaniem).

Ilustracją rozważań teoretycznych są badania symulacyjne przeprowadzone dla wybranych rozkładów. W przypadku próby o nieparzystej liczbie obserwacji wyznaczone zostały granice przedziałów ufności dokładną metodą percentyli oraz zbadany został wpływ liczebności próby oraz założonego poziomu ufności na dokładność oszacowania. Dla prób o parzystej liczbie obserwacji dodatkowo porównane zostaną wyniki estymacji przedziałowej uzyskane dwoma estymatorami – standardowym i w postaci mediany z próby.

2. WYBRANE ESTYMATORY MEDIANY

Rozważania prowadzone będą dla zmiennej losowej X określonej dla populacji o nieznanym rozkładzie F o medianie $M_{0,5}$. Z populacji pobrano prostą próbę losową $(X_1, X_2,\ldots, X_n)$, której niemalejąco uporządkowana realizacja oznaczona zostanie jako $(x_{(1)}, x_{(2)},\ldots, x_{(n)})$. Statystyka pozycyjna o randze k oznaczona będzie jako $X_{(k)}$.

Przyjmujemy, że medianą z próby o nieparzystej liczbie obserwacji jest statystyka:

$$Me = X_{\left(\frac{n+1}{2}\right)}, \tag{1}$$

zaś z próby o parzystej liczbie obserwacji statystyka:

$$Me = \frac{1}{2}\left(X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)} \right). \tag{2}$$

Mediana z próby jest medianowo-nieobciążonym estymatorem mediany dla *n* nieparzystego. Jak udowodnił Zieliński (1995), w przypadku parzystego *n*, mediana próby może znajdować się dowolnie daleko od szacowanej mediany rozkładu, ponieważ dla każdej stałej *C* istnieje rozkład *F*, dla którego odległość oszacowania od faktycznej wartości mediany będzie większa od *C*. Z twierdzenia tego nie wynika jednak, że ma to miejsce dla każdego rozkładu.

Z drugiej strony wiadomo (Domański, Pruska, 2000), że granicznym rozkładem mediany z próby wylosowanej z populacji o rozkładzie z gęstością *f* taką, że $f(M_{0,5}) > 0$ jest rozkład normalny:

$$Me \sim \mathrm{N}\left(M_{0,5}, \frac{1}{2\sqrt{n} f\left(M_{0,5}\right)} \right), \tag{3}$$

gdzie *f* jest funkcją gęstości badanej zmiennej losowej w populacji. Oznacza to, że mediana z próby jest asymptotycznie nieobciążonym estymatorem mediany populacji.

Dla *n* parzystego medianowo-nieobciążonym estymatorem mediany jest natomiast estymator standardowy (Pekasiewicz, 2015) określony jako:

$$Me^{st} = X_{\left(\frac{n}{2}+I_{(0,5)}(u)\right)}, \tag{4}$$

gdzie *u* jest wartością zmiennej losowej o rozkładzie jednostajnym na przedziale [0,1], a $I_{(0,5)}(u)$ indykatorem przyjmującym wartość 1 lub 0, w zależności od tego, czy *u* jest większe bądź równe 0,5, czy mniejsze.

3. ROZKŁADY BOOTSTRAPOWYCH ESTYMATORÓW MEDIANY

W metodzie bootstrapowej, z prostej próby pierwotnej losowane są ze zwracaniem próby wtórne, które oznaczone zostaną jako $\left(X_1^*, X_2^*, ..., X_n^*\right)$. Uporządkowany zbiór możliwych realizacji każdej ze zmiennych X_i^* to $\{x_{(1)}, x_{(2)}, \ldots, x_{(n)}\}$. Jeżeli w próbie pierwotnej nie występują powtórzenia, prawdopodobieństwo wylosowania każdej

wartości $x_{(i)}$ jest jednakowe i równe $\frac{1}{n}$. Statystyka pozycyjna próby wtórnej o randze k oznaczona będzie jako $X^*_{(k)}$.

Dla każdej próby wtórnej można wyznaczyć medianę według reguły (1) dla n nieparzystego oraz reguł (2) lub (4) dla n parzystego. Bootstrapowymi realizacjami estymatora mediany, w przypadku wykorzystania formuł (1) i (4), mogą być jedynie elementy próby pierwotnej. Dla estymatora określonego formułą (2) mogą być to nie tylko elementy próby pierwotnej, ale również średnie arytmetyczne wszystkich dwuelementowych kombinacji z niej wylosowanych.

3.1. PRÓBA O NIEPARZYSTEJ LICZBIE OBSERWACJI

W przypadku próby o nieparzystej liczbie elementów bootstrapowym estymatorem mediany jest statystyka pozycyjna $X^*_{(m)}$ o randze $m = \frac{n+1}{2}$. Rozkład tego estymatora jest określony formułą (por. Efron, 1979; Pekasiewicz, 2015):

$$\begin{cases} \text{dla } l = 2,3,...,n-1: \\ P\left(X^*_{(m)} = x_{(l)}\right) = \sum_{j=0}^{m-1}\left(\frac{l-1}{n}\right)^j\left(1-\frac{l-1}{n}\right)^{n-j}\binom{n}{j} - \sum_{j=0}^{m-1}\left(\frac{l}{n}\right)^j\left(1-\frac{l}{n}\right)^{n-j}\binom{n}{j}, \\ \text{dla } l = 1, n: \\ P\left(X^*_{(m)} = x_{(l)}\right) = \sum_{j=m}^{n}\left(\frac{1}{n}\right)^j\left(\frac{n-1}{n}\right)^{n-j}\binom{n}{j}. \end{cases} \tag{5}$$

Rozkład ten można przedstawić również następująco (na podstawie Evans i inni, 2006):

$$\begin{cases} \text{dla } l = 2,3,\ldots,n-1: \\ P\left(X^*_{(m)} = x_{(l)}\right) = \sum_{i=0}^{m-1}\sum_{j=0}^{n-m}\left(\frac{l-1}{n}\right)^i\left(\frac{1}{n}\right)^{n-i-j}\left(\frac{n-l}{n}\right)^j\binom{n}{i,j,n-i-j}, \\ \text{dla } l = 1, n: \\ P\left(X^*_{(m)} = x_{(l)}\right) = \sum_{j=m}^{n}\left(\frac{1}{n}\right)^j\left(\frac{n-1}{n}\right)^{n-j}\binom{n}{j}. \end{cases} \tag{6}$$

Rozkład określony formułą (5) lub (6) ma dwie istotne własności. Po pierwsze prawdopodobieństwa $P\left(X^*_{(m)} = x_{(l)}\right)$ nie zależą od wartości obserwacji wchodzących w skład próby, lecz jedynie od jej rozmiaru, a wobec tego są jednakowe dla wszystkich

n elementowych prób pierwotnych. Z własności tej wynika możliwość stworzenia gotowych tablic zawierających rozkłady estymatora dla różnych n[2]. Można ponadto pokazać, że dla $l = 1, 2,\ldots, n$:

$$P\left(X^*_{(m)} = x_{(l)}\right) = P\left(X^*_{(m)} = x_{(n-l+1)}\right). \tag{7}$$

3.2. PRÓBA O PARZYSTEJ LICZBIE OBSERWACJI

Rozważania prowadzone będą dla dwóch bootstrapowych estymatorów mediany. Pierwszy z nich określony regułą (2), oznaczony zostanie jako $X^{*a}_{(m_1,m_1+1)}$, drugi natomiast dany regułą (4) jako $X^{*b}_{(m_1,m_1+1)}$, gdzie $m_1 = \frac{n}{2}$.

Jak wspomniano wcześniej, w przypadku estymatora $X^{*a}_{(m_1,m_1+1)}$, medianą w próbach wtórnych mogą być wszystkie elementy próby oraz średnie arytmetyczne wszystkich kombinacji dwuelementowych z próby. Rozważmy dwie próby czteroelementowe: (1; 2; 3; 4) oraz (1; 2; 4; 7). W przypadku próby pierwszej medianą w próbach wtórnych mogą być wartości 1; 1,5; 2; 2,5; 3; 3,5; 4 natomiast dla próby drugiej 1; 1,5; 2; 2,5; 3, 4; 4,5; 5; 6; 8. Bootstrapowy estymator mediany $X^{*a}_{(m_1,m_1+1)}$ dla próby pierwszej ma 7 realizacji, zaś dla drugiej 10. Prawdopodobieństwa poszczególnych realizacji mediany będą więc różne dla różnych prób, ponieważ średnie arytmetyczne różnych kombinacji mogą być takie same, a ponadto mogą być równe elementom próby pierwotnej.

Następny problem wynika z kolejności możliwych wartości mediany w próbach wtórnych. Dla próby (1; 2; 2,4; 5) średnia arytmetyczna elementu pierwszego i trzeciego jest mniejsza od elementu drugiego, natomiast dla próby (1; 2; 4; 5) jest od niego większa.

Przedstawione problemy powodują, że nie jest możliwe podanie ogólnej formuły na prawdopodobieństwa dla poszczególnych realizacji estymatora. Można natomiast wyznaczyć prawdopodobieństwa, że na pozycjach m_1 i m_1+1 w próbach wtórnych występuje dowolny *l*-ty element próby, lub dwa dowolne jej elementy: l_1 i l_2, przy czym $l_1 < l_2$.

Prawdopodobieństwo, że w uporządkowanej próbie wtórnej *l*-ty element próby pierwotnej występuje na co najmniej dwóch środowych pozycjach m_1 i m_1+1 jest następujące:

---

[2] Na stronie http://joanna_kisielinska.users.sggw.pl podano tablice rozkładu bootstrapowego estymatora mediany dla liczebności nieparzystych z przedziału 3–201.

$$\begin{cases} \text{dla } l = 2,\ldots,n-1 \\ P\left(\left(X^*_{(m_1)} = x_{(l)}\right) \wedge \left(X^*_{(m_1+1)} = x_{(l)}\right)\right) = \sum\limits_{i=0}^{m_1-1}\sum\limits_{j=0}^{m_1-1}\left(\frac{l-1}{n}\right)^i\left(\frac{1}{n}\right)^{n-i-j}\left(\frac{n-l}{n}\right)^j\binom{n}{i,n-i-j,j}, \\ \text{dla } l = 1, n \\ P\left(\left(X^*_{(m_1)} = x_{(l)}\right) \wedge \left(X^*_{(m_1+1)} = x_{(l)}\right)\right) = \sum\limits_{i=m_1+1}^{n}\left(\frac{1}{n}\right)^i\left(\frac{n-1}{n}\right)^{n-i}\binom{n}{i}. \end{cases} \quad (8)$$

Można pokazać, że

$$P\left(\left(X^*_{(m_1)} = x_{(l)}\right) \wedge \left(X^*_{(m_1+1)} = x_{(l)}\right)\right) = P\left(\left(X^*_{(m_1)} = x_{(n-l+1)}\right) \wedge \left(X^*_{(m_1+1)} = x_{(n-l+1)}\right)\right).$$

Prawdopodobieństwo, że w uporządkowanej próbie wtórnej pierwszy element uporządkowanej próby pierwotnej występuje dokładnie m_1 razy, a element l-ty występuje przynajmniej raz na pozycji m_1+1 jest równe:

$$P\left(\left(X^*_{(m_1)} = x_{(1)}\right) \wedge \left(X^*_{(m_1+1)} = x_{(l)}\right)\right) = \sum_{i=0}^{m_1-1}\left(\frac{1}{n}\right)^{n-i}\left(\frac{n-l}{n}\right)^i\binom{n}{m_1, n-m_1-i, i}, \quad (9)$$

dla l = 2, 3, ..., n − 1.

Prawdopodobieństwo, że w uporządkowanej próbie wtórnej n-ty element uporządkowanej próby pierwotnej występuje dokładnie m_1 razy, a element l-ty występuje przynajmniej raz na pozycji m_1 jest równe:

$$P\left(\left(X^*_{(m_1)} = x_{(l)}\right) \wedge \left(X^*_{(m_1+1)} = x_{(n)}\right)\right) = \sum_{i=0}^{m_1-1}\left(\frac{l-1}{n}\right)^i\left(\frac{1}{n}\right)^{n-i}\binom{n}{i, n-m_1-i, m_1}, \quad (10)$$

przy czym $P\left(\left(X^*_{(m_1)} = x_{(1)}\right) \wedge \left(X^*_{(m_1+1)} = x_{(l)}\right)\right) = P\left(\left(X^*_{(m_1)} = x_{(n-l+1)}\right) \wedge \left(X^*_{(m_1+1)} = x_{(n)}\right)\right)$, dla l = 2, 3, …, n − 1.

Prawdopodobieństwo, że w próbie wtórnej pierwszy oraz n-ty element uporządkowanej próby pierwotnej występują dokładnie po m_1 razy jest równe:

$$P\left(X^*_{(m_1)} = x_{(1)}\right) = P\left(X^*_{(m_1+1)} = x_{(n)}\right) = \left(\frac{1}{n}\right)^n\binom{n}{m_1}. \quad (11)$$

Prawdopodobieństwo, że w uporządkowanej próbie wtórnej element l_1 znajduje się przynajmniej raz na pozycji m_1, a element l_2 przynajmniej raz na pozycji m_1+1 jest równe:

$$P\left(\left(X^*_{(m_1)} = x_{(l_1)}\right) \wedge \left(X^*_{(m_1+1)} = x_{(l_2)}\right)\right) =$$
$$= \sum_{i=0}^{m_1-1}\sum_{j=0}^{m_1-1}\left(\frac{l_1-1}{n}\right)^i\left(\frac{1}{n}\right)^{n-i-j}\left(\frac{n-l_2}{n}\right)^j\binom{n}{i, m_1-i, m_1-j, j}, \tag{12}$$
$$\forall\,(l_1,l_2): l_1 < l_2 \in \{2,3,...,n-2\}\times\{3,4,...,n-1\},$$

przy czym

$$P\left(\left(X^*_{(m_1)} = x_{(l_1)}\right) \wedge \left(X^*_{(m_1+1)} = x_{(l_2)}\right)\right) = P\left(\left(X^*_{(m_1)} = x_{(n-l_2+1)}\right) \wedge \left(X^*_{(m_1+1)} = x_{(n-l_1+1)}\right)\right).$$

Aby wyznaczyć rozkład estymatora $X^{*a}_{(m_1,m_1+1)}$ należy określić wszystkie możliwe jego realizacje i je uporządkować, a następnie ze wzorów od (8) do (12) obliczyć prawdopodobieństwa ich wystąpienia[3]. Prawdopodobieństwa te należy w pewnych przypadkach zsumować. Ponieważ nie jest z góry znana liczba oraz kolejność wystąpienia uporządkowanych realizacji estymatora $X^{*a}_{(m_1,m_1+1)}$, nie można podać ogólnej postaci jego rozkładu – prawdopodobieństwa dla poszczególnych realizacji estymatora zależą od elementów próby.

Realizacjami estymatora $X^{*b}_{(m_1,m_1+1)}$ są jedynie elementy próby. Prawdopodobieństwa dane wzorami od (8) do (12) należy podzielić na pół i przypisać je elementom występującym na pozycji m_1 i m_1+1. Ostatecznie prawdopodobieństwa dla poszczególnych realizacji estymatora określone są następująco:

$$\begin{cases} P\left(X^{*b}_{(m_1,m_1+1)} = x_{(1)}\right) = P\left(X^*_{(m_1,m_1+1)} = x_{(n)}\right) = \sum_{i=m_1+1}^{n}\left(\frac{1}{n}\right)^i\left(\frac{n-1}{n}\right)^{n-i}\binom{n}{i} + \\ +\frac{1}{2}\sum_{l=2}^{n-1}\sum_{i=0}^{m_1-1}\left(\frac{1}{n}\right)^{n-i}\left(\frac{n-l}{n}\right)^i\binom{n}{m_1, n-m_1-i, i} + \frac{1}{2}\left(\frac{1}{n}\right)^n\binom{n}{m_1}, \end{cases} \tag{13}$$

$$\begin{cases} P\left(X^{*b}_{(m_1,m_1+1)} = x_{(2)}\right) = P\left(X^*_{(m_1,m_1+1)} = x_{(n-1)}\right) = \sum_{i=0}^{m_1-1}\sum_{j=0}^{m_1-1}\left(\frac{1}{n}\right)^{n-j}\left(\frac{n-2}{n}\right)^j\binom{n}{i, n-i-j, j} + \\ +\frac{1}{2}\sum_{i=0}^{m_1-1}\left(\frac{1}{n}\right)^{n-i}\left(\frac{n-2}{n}\right)^i\binom{n}{m_1, n-m_1-i, i} + \frac{1}{2}\sum_{i=0}^{m_1-1}\left(\frac{1}{n}\right)^n\binom{n}{i, n-m_1-i, m_1} + \\ +\frac{1}{2}\sum_{l=3}^{n-1}\sum_{i=0}^{m_1-1}\sum_{j=0}^{m_1-1}\left(\frac{1}{n}\right)^{n-j}\left(\frac{n-l}{n}\right)^j\binom{n}{i, m_1-i, m_1-j, j}, \end{cases} \tag{14}$$

[3] Prawdopodobieństwa te umieszczono na stronie http://joanna_kisielinska.users.sggw.pl dla liczebności parzystych z przedziału 4–200.

$$\begin{cases} P\left(X^{*b}_{(m_1,m_1+1)}=x_{(l)}\right)=P\left(X^{*b}_{(m_1,m_1+1)}=x_{(n-l+1)}\right)= \\ =\sum_{i=0}^{m_1-1}\sum_{j=0}^{m_1-1}\left(\frac{l-1}{n}\right)^i\left(\frac{1}{n}\right)^{n-i-j}\left(\frac{n-l}{n}\right)^j\binom{n}{i,n-i-j,j}+ \\ +\frac{1}{2}\sum_{i=0}^{m_1-1}\left(\frac{1}{n}\right)^{n-i}\left(\frac{n-l}{n}\right)^i\binom{n}{m_1,n-m_1-i,i}+\frac{1}{2}\sum_{i=0}^{m_1-1}\left(\frac{l-1}{n}\right)^i\left(\frac{1}{n}\right)^{n-i}\binom{n}{i,n-m_1-i,m_1}+ \\ +\frac{1}{2}\sum_{l_1=2}^{l-1}\sum_{i=0}^{m_1-1}\sum_{j=0}^{m_1-1}\left(\frac{l_1-1}{n}\right)^i\left(\frac{1}{n}\right)^{n-i-j}\left(\frac{n-l}{n}\right)^j\binom{n}{i,m_1-i,m_1-j,j}+ \\ +\frac{1}{2}\sum_{l_2=l+1}^{n-1}\sum_{i=0}^{m_1-1}\sum_{j=0}^{m_1-1}\left(\frac{l-1}{n}\right)^i\left(\frac{1}{n}\right)^{n-i-j}\left(\frac{n-l_2}{n}\right)^j\binom{n}{i,m_1-i,m_1-j,j}, \quad \text{dla } l=3,4,\ldots,m_1. \end{cases} \tag{15}$$

Rozkład dany formułami od (13) do (15) ma dwie istotne własności, które cechowały rozkład estymatora mediany dla próby o nieparzystej liczbie obserwacji. Tak więc $P\left(X^{*b}_{(m_1,m_1+1)}=x_{(l)}\right)$, dla $l = 1, 2, \ldots, n$ są jednakowe dla wszystkich prób o zadanej liczebności (co umożliwia opracowanie gotowych tablic[4]) oraz spełniają warunek (7).

4. SZACOWANIE MEDIANY DOKŁADNĄ METODĄ PERCENTYLI

Do wyznaczenia przedziałów ufności dla mediany można użyć metody percentyli, opisanej miedzy innymi w pracach: Domański, Pruska (2000), Wilcox (2001), Pekasiewicz (2015). Podkreślić należy, że losowanie wtórnych prób bootstrapowych z całej ich populacji (o liczebności n^n) nie jest tu potrzebne. Dla estymatorów $X^*_{(m)}$ i $X^{*b}_{(m_1,m_1+1)}$ można z góry określić, które elementy próby (pierwotnej) są granicami przedziałów dla zadanych poziomów ufności, ponieważ prawdopodobieństwa poszczególnych ich realizacji nie zależą od elementów próby, a jedynie od jej rozmiaru. W przypadku estymatora $X^{*a}_{(m_1,m_1+1)}$ wyznaczenie granic wymaga dodatkowych obliczeń, związanych z wyznaczaniem rozkładu (prawdopodobieństwa poszczególnych jego realizacji zależą od elementów próby) i w tym przypadku losowanie prób wtórnych może być uzasadnione, aczkolwiek niekonieczne.

Przewaga dokładnej metody percentyli nad metodą klasyczną (z losowaniem) dla estymatorów $X^*_{(m)}$ i $X^{*b}_{(m_1,m_1+1)}$ wynika przede wszystkim z jej prostoty. Wiadomo, że stosując dokładną metoda bootstrapową nie wprowadza się dodatkowych błędów, wynikających z wtórnego próbkowania (Hutson, Ernst, 2000). W przypadku metody klasycznej liczba losowanych prób wtórnych jest duża, a wobec tego ewentualne błędy będą niewielkie.

Wyznaczając granice przedziałów ufności dla mediany należy pamiętać, że rozkłady jej bootstrapowych estymatorów dla skończonego n są dyskretne, a wobec tego

[4] Na stronie http://joanna_kisielinska.users.sggw.pl podano tablice rozkładu bootstrapowego standardowego estymatora mediany dla liczebności parzystych z przedziału 4–200.

ich dystrybuanty nie przyjmują wszystkich wartości z przedziału [0,1]. Dla zadanego poziomu ufności $1-\alpha$, lewą granicą przedziału ufności mediany będzie największa realizacja estymatora spełniająca warunek:

$$F\left(x^*_{(d)}\right) \leq \frac{\alpha}{2}, \tag{16}$$

gdzie F jest dystrybuantą wybranego estymatora mediany. Granicą prawą przedziału jest natomiast realizacja najmniejsza, taka że:

$$F\left(x^*_{(g)}\right) \geq 1-\frac{\alpha}{2}. \tag{17}$$

Brak równości we wzorach (16) i (17), wynikający z dyskretności rozkładu estymatora, powoduje, że rzeczywisty poziom istotności, który oznaczony zostanie jako α_f, będzie zwykle mniejszy od założonego α.

4.1. PRÓBA O NIEPARZYSTEJ LICZBIE OBSERWACJI

Jak wspomniano, realizacjami estymatora $X^*_{(m)}$ są jedynie elementy próby pierwotnej, wobec czego we wzorach (16) i (17) zamiast $x^*_{(d)}$ i $x^*_{(g)}$ należałoby wstawić $x_{(d)}$ i $x_{(g)}$. Ponieważ jego rozkład określony formułami (5) lub (6) nie zależy od elementów próby, a jedynie od jej liczebności, pozycje d i g określające granice przedziałów ufności można z góry określić dla dowolnych prób o zadanej liczebności[5]. Ze względu na własność (7) zachodzi ponadto: $g = n - d + 1$.

Znając granice przedziałów ufności można wyznaczyć faktyczny poziom istotności, który spełnia warunki: $F\left(x^*_{(d)}\right) = \frac{\alpha_f}{2}$ oraz $F\left(x^*_{(g)}\right) = 1-\frac{\alpha_f}{2}$.

Połowa długości przedziału ufności jest miarą dokładności oszacowania mediany. Dla małych prób jest ona średnią odległością granic przedziału od mediany estymatora. Z własności (7) wynika bowiem, że granice przedziałów ufności nie będą równo oddalone od mediany estymatora, jeżeli próba nie spełnia następującego warunku:

$$\forall l \in [1, m-1]: x_{(m)} - x_{(l)} = x_{(n-l+1)} - x_{(m)}, \tag{18}$$

Dla prób dużych, miarą dokładności może być natomiast wariancja lub odchylenie standardowe estymatora, ponieważ wartość oczekiwana staje się wówczas bliska medianie (rozkładem granicznym mediany jest rozkład normalny dany formułą (3)). Wariancję lub odchylenie standardowe bootstrapowego estymatora mediany można wyznaczyć, ponieważ znany jest jego rozkład (określony formułą (5 lub 6). Można także wykorzystać wzory podane w pracy Maritz, Jarrett (1978).

---

[5] Numery obserwacji stanowiące granice przedziałów ufności dla liczebności nieparzystych z przedziału 3–201 i wybranych poziomów ufności, podano na stronie http://joanna_kisielinska.users.sggw.pl.

4.2. PRÓBA O PARZYSTEJ LICZBIE OBSERWACJI

Jak pokazano, nie można podać ogólnej postaci rozkładu estymatora $X^{*a}_{(m_1,m_1+1)}$, ponieważ zależy on od elementów próby. W konsekwencji tego nie jest możliwe określenie z góry granic przedziałów ufności. Dla każdej małej próby trzeba wyznaczać je indywidualnie. Ponieważ prawdopodobieństwo odpowiadające dowolnej *l*-tej realizacji estymatora $X^{*a}_{(m_1,m_1+1)}$ może być różne od prawdopodobieństwa dla realizacji o pozycji $n - l + 1$, prawdopodobieństwa $\alpha_{fd} = P\left(X^{*a}_{(m_1,m_1+1)} < x^{*}_{(d)}\right)$ i $\alpha_{fg} = P\left(X^{*a}_{(m_1,m_1+1)} > x^{*}_{(g)}\right)$ mogą być różne. Prawdopodobieństwo α_{fd} jest faktycznym lewostronnym poziomem istotności, zaś α_{fg} prawostronnym.

Z własności estymatora $X^{*b}_{(m_1,m_1+1)}$ (tak jak dla estymatora $X^{*}_{(m)}$) wynika, że można z góry określić pozycje granic przedziałów ufności dla wszystkich prób o zadanej liczebności[6], przy czym między numerami obserwacji stanowiącymi jego granice zachodzi zależność:

$$g = n - d + 1.$$

5. WYNIKI BADAŃ

Badania symulacyjne przeprowadzono dla populacji o wybranych sześciu rozkładach – czterech asymetrycznych i dwóch symetrycznych. Rozkłady te oznaczone zostały jako *F*(20,10), *F*(2,10), *CHI*(2), *CHI*(10), *N*(5,1) i *N*(5,4).

Dla różnych liczebności losowano 1000 razy po *n* liczb pseudolosowych z przedziału [0,1], które traktowano jako dystrybuantę rozkładów. Zastosowano jednakowe wartości dystrybuant dla wszystkich rozkładów, co pozwala uzyskać lepszą porównywalność wyników. Dobór taki uniezależnia wyniki obliczeń dla poszczególnych rozkładów od jakości generatora liczb pseudolosowych. Mając wylosowanych *n* wartości dystrybuanty, dla każdego z wziętego pod uwagę rozkładów, określano próby losowe. Dla 1000 wylosowanych w ten sposób prób wyznaczono rozkłady bootstrapowych estymatorów mediany, co pozwoliło na jej przedziałowe oszacowanie.

Wszystkie obliczenia wykonane zostały w Excelu z wykorzystaniem języka VBA for Application.

5.1. ESTYMACJA MEDIANY DOKŁADNĄ METODĄ BOOTSTRAPOWĄ DLA PRÓBY O NIEPARZYSTEJ LICZBIE OBSERWACJI

Badania przeprowadzono dla prób o liczebnościach od 11 do 201. Liczebności dobrano tak, aby obejmowały zarówno małe jak i duże próby. Aby możliwa była prezentacja wyników liczebności zwiększano o 10.

---

[6] Numery obserwacji stanowiące granice przedziałów ufności dla liczebności parzystych z przedziału 4–200 i wybranych poziomów ufności, podano na stronie http://joanna_kisielinska.users.sggw.pl.

Wyniki estymacji przedziałowej mediany dokładną metodą percentyli dla wybranych rozkładów przy użyciu estymatora $X^*_{(m)}$ przedstawiono w tabelach 1 i 2.

Tabela 1.

Lewa granica przedziału ufności oraz faktyczny i zliczeniowy poziom istotności uzyskany dokładną metodą percentyli dla *n* nieparzystego

n	α = 0,05			α = 0,04			α = 0,03			α = 0,02			α = 0,01		
	d	α_f	α_o	*d*	α_f	α_o	*d*	α_f	α_o	*d*	α_f	α_o	*d*	α_f	α_o
11	2	0,014	0,012	2	0,014	0,012	2	0,014	0,012	2	0,014	0,012	1	0,000	0,000
21	6	0,037	0,031	6	0,037	0,031	5	0,009	0,004	5	0,009	0,004	5	0,009	0,004
31	10	0,039	0,039	10	0,039	0,039	9	0,014	0,019	9	0,014	0,019	8	0,004	0,011
41	14	0,036	0,024	14	0,036	0,024	13	0,015	0,016	13	0,015	0,016	12	0,005	0,012
51	18	0,031	0,019	18	0,031	0,019	17	0,014	0,004	17	0,014	0,004	16	0,006	0,000
61	22	0,026	0,016	22	0,026	0,016	22	0,026	0,016	21	0,012	0,012	20	0,005	0,008
71	27	0,040	0,034	27	0,040	0,034	26	0,021	0,024	25	0,011	0,016	24	0,005	0,008
81	31	0,032	0,027	31	0,032	0,027	30	0,017	0,021	30	0,017	0,021	29	0,009	0,018
91	36	0,043	0,034	35	0,025	0,022	35	0,025	0,022	34	0,014	0,014	33	0,007	0,007
101	40	0,034	0,030	40	0,034	0,030	39	0,020	0,015	38	0,011	0,011	37	0,006	0,008
111	45	0,044	0,020	44	0,027	0,017	44	0,027	0,017	43	0,016	0,011	42	0,009	0,006
121	49	0,034	0,034	49	0,034	0,034	48	0,021	0,026	47	0,013	0,013	46	0,007	0,009
131	54	0,042	0,028	53	0,027	0,014	53	0,027	0,014	52	0,017	0,003	50	0,006	0,001
141	58	0,034	0,029	58	0,034	0,029	57	0,021	0,020	56	0,013	0,012	55	0,008	0,009
151	63	0,040	0,035	62	0,027	0,027	62	0,027	0,027	61	0,017	0,019	59	0,006	0,001
161	68	0,047	0,047	67	0,032	0,029	66	0,021	0,021	65	0,014	0,015	64	0,008	0,012
171	72	0,037	0,019	72	0,037	0,019	71	0,025	0,012	70	0,017	0,008	68	0,007	0,007
181	77	0,043	0,028	76	0,030	0,021	76	0,030	0,021	74	0,013	0,015	73	0,009	0,012
191	82	0,049	0,045	81	0,035	0,032	80	0,024	0,022	79	0,016	0,013	77	0,007	0,008
201	86	0,040	0,034	86	0,040	0,034	85	0,028	0,024	84	0,019	0,021	82	0,008	0,011

Uwagi: *d* – numer obserwacji, która jest dolną granicą przedziału ufności, α_f – faktyczny poziom ufności, α_o – udział przedziałów niezawierających mediany rozkładu wyznaczony z 1000 prób.

Źródło: badania własne.

W tabeli 1 przedstawiono numery obserwacji stanowiące lewą granice przedziałów ufności oraz faktyczne poziomy istotności α_f. Mając numer obserwacji stanowiący lewą granicę przedziału, można dla dowolnej próby losowej natychmiast (bez konieczności losowania prób wtórnych) wyznaczyć przedziały ufności. Ponieważ rozkład estymatora

jest dyskretny, faktyczne poziomy ufności $1-\alpha_f$ są wyższe od założonych. W tabeli 1 podano ponadto udziały przedziałów niezawierających mediany rozkładu α_o (zliczeniowe poziomy istotności). Udziały te były zwykle mniejsze od α_f. Ze względu na sposób losowania prób udziały przedziałów niezawierających mediany były jednakowe dla wszystkich rozkładów. Wartości te nie stanowią dokładnej informacji o faktycznych poziomach ufności, można je jedynie traktować jako sprawdziany poprawności działania zastosowanego generatora liczb pseudolosowych.

Tabela 2.

Średnie szerokości połowy przedziałów ufności dla $\alpha = 0{,}05$ oraz średnie oszacowania odchylenia standardowego wyznaczone z 1000 prób dla n nieparzystego

n	F(20,10) $M_{0,5} = 1{,}0349$ $\sigma = 0{,}8539$		F(2,10) $M_{0,5} = 0{,}7435$ $\sigma = 1{,}6137$		CHI(2) $M_{0,5} = 1{,}3863$ $\sigma = 2{,}0000$		CHI(10) $M_{0,5} = 9{,}3418$ $\sigma = 4{,}4721$		N(5,1) $M_{0,5} = 5$ $\sigma = 1{,}000$		N(5,4) $M_{0,5} = 5$ $\sigma = 4{,}0000$	
	d	σ_b	d	σ_b	d	σ_b	d	σ_b	d	σ_b	d	σ_b
11	0,6997	0,2657	1,1761	0,4575	1,8108	0,7132	4,5660	1,7577	1,0511	0,4017	4,2045	1,6067
21	0,3814	0,1764	0,6177	0,2909	1,0304	0,4833	2,7045	1,2451	0,6265	0,2871	2,5060	1,1484
31	0,2929	0,1444	0,4705	0,2355	0,7964	0,3966	2,1108	1,0347	0,4902	0,2394	1,9610	0,9576
41	0,2592	0,1238	0,4142	0,2000	0,7071	0,3402	1,8851	0,8966	0,4385	0,2080	1,7541	0,8320
51	0,2400	0,1129	0,3831	0,1820	0,6556	0,3107	1,7491	0,8213	0,4068	0,1906	1,6271	0,7624
61	0,2229	0,1007	0,3562	0,1620	0,6103	0,2773	1,6252	0,7334	0,3773	0,1700	1,5090	0,6800
71	0,1898	0,0927	0,3026	0,1490	0,5205	0,2554	1,3899	0,6759	0,3229	0,1566	1,2915	0,6264
81	0,1826	0,0855	0,2904	0,1367	0,5005	0,2353	1,3408	0,6272	0,3120	0,1458	1,2480	0,5833
91	0,1621	0,0808	0,2574	0,1291	0,4443	0,2224	1,1927	0,5934	0,2778	0,1380	1,1112	0,5522
101	0,1631	0,0774	0,2594	0,1237	0,4477	0,2131	1,1991	0,5675	0,2788	0,1318	1,1153	0,5272
111	0,1461	0,0729	0,2322	0,1163	0,4014	0,2008	1,0760	0,5358	0,2502	0,1245	1,0010	0,4981
121	0,1466	0,0698	0,2330	0,1114	0,4028	0,1923	1,0793	0,5128	0,2510	0,1191	1,0039	0,4764
131	0,1349	0,0668	0,2142	0,1065	0,3706	0,1840	0,9940	0,4915	0,2312	0,1142	0,9249	0,4569
141	0,1351	0,0638	0,2146	0,1018	0,3713	0,1759	0,9952	0,4695	0,2313	0,1090	0,9254	0,4361
151	0,1274	0,0625	0,2023	0,0995	0,3503	0,1722	0,9401	0,4605	0,2187	0,1070	0,8748	0,4282
161	0,1184	0,0597	0,1878	0,0950	0,3254	0,1644	0,8741	0,4402	0,2034	0,1024	0,8136	0,4096
171	0,1213	0,0587	0,1925	0,0934	0,3335	0,1616	0,8945	0,4325	0,2080	0,1006	0,8320	0,4022
181	0,1133	0,0564	0,1797	0,0898	0,3115	0,1555	0,8369	0,4165	0,1948	0,0969	0,7790	0,3876
191	0,1066	0,0546	0,1691	0,0867	0,2933	0,1503	0,7884	0,4036	0,1835	0,0940	0,7341	0,3759
201	0,1086	0,0534	0,1722	0,0848	0,2986	0,1470	0,8021	0,3939	0,1866	0,0916	0,7465	0,3665

Uwagi: d – średnie szerokości połowy przedziałów ufności dla $\alpha = 0{,}05$, σ_b – średnie oszacowania odchylenia standardowego estymatora bootstrapowego.

Źródło: badania własne.

Tabela 2 zawiera średnie szerokości połowy przedziałów ufności dla $\alpha = 0{,}05$ oraz średnie oszacowania odchylenia standardowego estymatora wyznaczone z 1000 prób dla wybranych rozkładów. Oczywiste jest, że wraz ze wzrostem liczebności próby obydwie miary dokładności szacunku zmniejszają się, czyli oszacowania są coraz bardziej precyzyjne. Zauważyć należy, że z wyjątkiem $n = 11$ połowy długości przedziałów ufności dla $\alpha = 0{,}05$ są mniej więcej dwukrotnie większe niż odchylenie standardowe. Uzyskane wyniki nie stoją więc w sprzeczności z wynikami Wilcoxa (2001), na które powołuje się Pekasiewicz (2015) (str. 53), że rozkład mediany z próby jest zbieżny do rozkładu normalnego dla prób już 20-elementowych.

5.2. ESTYMACJA MEDIANY DOKŁADNĄ METODĄ BOOTSTRAPOWĄ DLA PRÓBY O PARZYSTEJ LICZBIE OBSERWACJI

Badania wykonano dla prób o liczebnościach od 10 do 200 (zwiększając liczebności o 10). Symulacje przeprowadzono używając estymatorów $X^{*a}_{(m_1,m_1+1)}$ i $X^{*b}_{(m_1,m_1+1)}$. Dla obydwu estymatorów zastosowano takie same zestawy 1000 prób losowych dla każdej z liczebności.

W tabelach 3, 4 i 5 przedstawiono wyniki estymacji przedziałowej mediany dokładną metodą percentyli dla wybranych rozkładów. Tabela 3 zawiera dla estymatora $X^{*a}_{(m_1,m_1+1)}$ średnie faktyczne poziomy istotności $\overline{\alpha}_{fd}$ i $\overline{\alpha}_{fg}$ obliczone z 1000 prób. Pamiętać bowiem należy, że w ogólnym przypadku wartości α_{fd} i α_{fg} są różne dla poszczególnych prób i rozkładów, choć badania pokazały, że różnice te były niewielkie. Bardzo małe są również różnice pomiędzy $\overline{\alpha}_{fd}$ i $\overline{\alpha}_{fg}$.

W przypadku estymatora $X^{*b}_{(m_1,m_1+1)}$ podano numery obserwacji będące lewymi granicami przedziałów ufności, oraz połowę faktycznego poziomu istotności $\frac{\alpha_f}{2}$. Porównując $\frac{\alpha_f}{2}$ z $\overline{\alpha}_{fd}$ i $\overline{\alpha}_{fg}$ stwierdzić można, że z wyjątkiem nielicznych przypadków $\frac{\alpha_f}{2}$ jest mniejsze od $\overline{\alpha}_{fd}$ i $\overline{\alpha}_{fg}$, co jest konsekwencją znacznie mniejszej liczby realizacji estymatora $X^{*b}_{(m_1,m_1+1)}$ od $X^{*a}_{(m_1,m_1+1)}$. Na podkreślenie zasługuje fakt, że różnice pomiędzy $\frac{\alpha_f}{2}$ i $\overline{\alpha}_{fd}$ oraz $\overline{\alpha}_{fg}$ są zwykle małe.

W tabeli 4 podano średnie szerokości połowy przedziałów ufności dla $\alpha = 0{,}05$ oraz średnie oszacowania odchylenia standardowego estymatora wyznaczone z 1000 prób dla estymatora $X^{*a}_{(m_1,m_1+1)}$, zaś w tabeli 5 dla estymatora $X^{*b}_{(m_1,m_1+1)}$. Podkreślić należy, że połowy długości przedziałów ufności oraz odchylenia standardowe uzyskane estymatorem $X^{*a}_{(m_1,m_1+1)}$ są dla wszystkich badanych rozkładów i wszystkich liczebności prób mniejsze, niż uzyskane estymatorem $X^{*b}_{(m_1,m_1+1)}$. Można więc stwierdzić, że dla badanych rozkładów estymator $X^{*b}_{(m_1,m_1+1)}$ ma mniejszą wariancję i pozwala uzyskać dokładniejsze przedziałowe oszacowanie mediany niż estymator $X^{*a}_{(m_1,m_1+1)}$.

Tabela 3.

Faktyczne poziomy istotności dla estymatora $X^{*a}_{(m_1,m_1+1)}$ oraz lewa granica przedziału ufności dla estymatora $X^{*b}_{(m_1,m_1+1)}$ uzyskane dokładną metodą percentyli dla n parzystego

n	$\alpha = 0{,}05$				$\alpha = 0{,}04$				$\alpha = 0{,}03$				$\alpha = 0{,}02$				$\alpha = 0{,}01$			
	$X^{*a}_{(m_1,m_1+1)}$		$X^{*b}_{(m_1,m_1+1)}$		$X^{*a}_{(m_1,m_1+1)}$		$X^{*b}_{(m_1,m_1+1)}$		$X^{*a}_{(m_1,m_1+1)}$		$X^{*b}_{(m_1,m_1+1)}$		$X^{*a}_{(m_1,m_1+1)}$		$X^{*b}_{(m_1,m_1+1)}$		$X^{*a}_{(m_1,m_1+1)}$		$X^{*b}_{(m_1,m_1+1)}$	
	$\overline{\alpha}_{fd}$	$\overline{\alpha}_{fg}$	d	$\frac{\alpha_f}{2}$	$\overline{\alpha}_{fd}$	$\overline{\alpha}_{fg}$	d	$\frac{\alpha_f}{2}$	$\overline{\alpha}_{fd}$	$\overline{\alpha}_{fg}$	d	$\frac{\alpha_f}{2}$	$\overline{\alpha}_{fd}$	$\overline{\alpha}_{fg}$	d	$\frac{\alpha_f}{2}$	$\overline{\alpha}_{fd}$	$\overline{\alpha}_{fg}$	d	$\frac{\alpha_f}{2}$
10	0,0195	0,0195	2	0,0196	0,0195	0,0195	2	0,0196	0,0070	0,0070	1	0,0009	0,0070	0,0070	1	0,0009	0,0011	0,0012	1	0,0009
20	0,0208	0,0208	5	0,0089	0,0177	0,0171	5	0,0089	0,0109	0,0112	5	0,0089	0,0092	0,0092	5	0,0089	0,0046	0,0047	4	0,0016
30	0,0226	0,0226	9	0,0117	0,0171	0,0174	9	0,0117	0,0138	0,0136	9	0,0117	0,0077	0,0077	8	0,0035	0,0042	0,0043	8	0,0035
40	0,0206	0,0205	13	0,0115	0,0182	0,0183	13	0,0115	0,0132	0,0136	13	0,0115	0,0083	0,0083	12	0,0043	0,0045	0,0046	12	0,0043
50	0,0238	0,0239	18	0,0222	0,0175	0,0175	17	0,0104	0,0132	0,0130	17	0,0104	0,0078	0,0078	16	0,0044	0,0046	0,0047	16	0,0044
60	0,0222	0,0222	22	0,0181	0,0194	0,0194	22	0,0181	0,0141	0,0143	21	0,0089	0,0095	0,0096	21	0,0089	0,0045	0,0046	20	0,0041
70	0,0224	0,0224	26	0,0146	0,0181	0,0179	26	0,0146	0,0119	0,0119	26	0,0146	0,0085	0,0087	25	0,0075	0,0042	0,0042	24	0,0036
80	0,0234	0,0232	31	0,0211	0,0178	0,0177	30	0,0118	0,0132	0,0135	30	0,0118	0,0096	0,0096	29	0,0062	0,0045	0,0047	28	0,0031
90	0,0240	0,0239	35	0,0166	0,0186	0,0187	35	0,0166	0,0141	0,0141	34	0,0094	0,0090	0,0088	34	0,0094	0,0041	0,0042	32	0,0026
100	0,0234	0,0235	40	0,0219	0,0190	0,0190	39	0,0131	0,0140	0,0140	39	0,0131	0,0091	0,0089	38	0,0076	0,0046	0,0047	37	0,0042
110	0,0241	0,0242	44	0,0172	0,0187	0,0188	44	0,0172	0,0141	0,0143	43	0,0104	0,0089	0,0089	42	0,0061	0,0047	0,0047	41	0,0034
120	0,0238	0,0235	49	0,0216	0,0190	0,0189	48	0,0136	0,0145	0,0144	48	0,0136	0,0093	0,0094	47	0,0083	0,0044	0,0042	46	0,0048
130	0,0233	0,0233	53	0,0170	0,0188	0,0187	53	0,0170	0,0144	0,0143	52	0,0107	0,0094	0,0093	51	0,0066	0,0045	0,0045	50	0,0039
140	0,0233	0,0230	58	0,0207	0,0184	0,0184	57	0,0135	0,0143	0,0143	57	0,0135	0,0094	0,0094	56	0,0085	0,0045	0,0045	54	0,0031
150	0,0220	0,0220	63	0,0245	0,0187	0,0187	62	0,0164	0,0145	0,0145	61	0,0106	0,0094	0,0094	60	0,0067	0,0046	0,0046	59	0,0042
160	0,0232	0,0232	67	0,0194	0,0175	0,0175	67	0,0194	0,0142	0,0142	66	0,0130	0,0093	0,0093	65	0,0084	0,0047	0,0047	63	0,0033
170	0,0239	0,0239	72	0,0226	0,0187	0,0187	71	0,0154	0,0138	0,0138	70	0,0103	0,0091	0,0091	69	0,0067	0,0047	0,0047	68	0,0043
180	0,0235	0,0235	76	0,0180	0,0190	0,0191	76	0,0180	0,0140	0,0140	75	0,0123	0,0093	0,0093	74	0,0082	0,0047	0,0047	72	0,0034
190	0,0237	0,0237	81	0,0207	0,0188	0,0188	80	0,0144	0,0140	0,0141	80	0,0144	0,0088	0,0088	79	0,0098	0,0047	0,0047	77	0,0042
200	0,0243	0,0243	86	0,0234	0,0189	0,0189	85	0,0166	0,0145	0,0145	84	0,0115	0,0093	0,0093	83	0,0078	0,0046	0,0046	81	0,0034

Uwagi: $\overline{\alpha}_{fd}$ – średni faktyczny poziom istotności lewostronny, $\overline{\alpha}_{fg}$ – średni faktyczny poziom istotności prawostronny, d – numer obserwacji, która jest dolną granicą przedziału ufności, α_f – połowa faktycznego poziom istotności.

Źródło: badania własne.

Tabela 4.

Średnie szerokości połowy przedziałów ufności estymatora $X^{*a}_{(m_1,m_1+1)}$ dla $\alpha = 0{,}05$ oraz średnie oszacowania odchylenia standardowego wyznaczone z 1000 prób dla n parzystego

n	F(20,10) $M_{0,5} = 1{,}0349$ $\sigma = 0{,}8539$		F(2,10) $M_{0,5} = 0{,}7435$ $\sigma = 1{,}6137$		CHI(2) $M_{0,5} = 1{,}3863$ $\sigma = 2{,}0000$		CHI(10) $M_{0,5} = 9{,}3418$ $\sigma = 4{,}4721$		N(5,1) $M_{0,5} = 5$ $\sigma = 1{,}000$		N(5,4) $M_{0,5} = 5$ $\sigma = 4{,}0000$	
	d	σ_b	d	σ_b	d	σ_b	d	σ_b	d	σ_b	d	σ_b
10	0,5217	0,2763	0,8681	0,4813	1,3657	0,7292	3,4855	1,7866	0,8042	0,4087	3,2169	1,6348
20	0,3281	0,1842	0,5333	0,3046	0,8941	0,5050	2,3394	1,2951	0,5438	0,2977	2,1751	1,1908
30	0,2579	0,1469	0,4141	0,2401	0,7023	0,4033	1,8598	1,0487	0,4302	0,2422	1,7207	0,9688
40	0,2337	0,1220	0,3728	0,1966	0,6375	0,3345	1,6926	0,8858	0,3941	0,2063	1,5765	0,8253
50	0,2051	0,1102	0,3278	0,1776	0,5615	0,3032	1,4949	0,8009	0,3467	0,1857	1,3869	0,7430
60	0,1818	0,0962	0,2882	0,1536	0,4972	0,2641	1,3366	0,7060	0,3118	0,1646	1,2470	0,6586
70	0,1727	0,0927	0,2747	0,1483	0,4733	0,2550	1,2677	0,6798	0,2950	0,1582	1,1798	0,6329
80	0,1652	0,0860	0,2639	0,1387	0,4540	0,2376	1,2080	0,6248	0,2798	0,1440	1,1193	0,5761
90	0,1483	0,0797	0,2355	0,1271	0,4072	0,2194	1,0932	0,5866	0,2545	0,1365	1,0179	0,5460
100	0,1393	0,0742	0,2214	0,1187	0,3825	0,2044	1,0257	0,5441	0,2387	0,1264	0,9549	0,5055
110	0,1354	0,0747	0,2154	0,1195	0,3722	0,2060	0,9970	0,5479	0,2317	0,1271	0,9268	0,5083
120	0,1321	0,0692	0,2103	0,1102	0,3633	0,1905	0,9710	0,5099	0,2253	0,1187	0,9013	0,4748
130	0,1280	0,0679	0,2031	0,1082	0,3517	0,1869	0,9441	0,4998	0,2197	0,1162	0,8788	0,4650
140	0,1217	0,0639	0,1936	0,1019	0,3349	0,1760	0,8959	0,4695	0,2080	0,1090	0,8321	0,4360
150	0,1184	0,0622	0,1879	0,0991	0,3256	0,1714	0,8737	0,4581	0,2032	0,1065	0,8129	0,4259
160	0,1125	0,0600	0,1784	0,0955	0,3093	0,1653	0,8314	0,4427	0,1935	0,1030	0,7742	0,4119
170	0,1111	0,0584	0,1763	0,0930	0,3056	0,1609	0,8206	0,4307	0,1909	0,1001	0,7637	0,4005
180	0,1053	0,0554	0,1670	0,0881	0,2895	0,1526	0,7779	0,4085	0,1810	0,0950	0,7240	0,3799
190	0,1031	0,0547	0,1634	0,0870	0,2835	0,1507	0,7621	0,4039	0,1774	0,0939	0,7095	0,3757
200	0,1009	0,0533	0,1600	0,0847	0,2776	0,1467	0,7458	0,3931	0,1735	0,0914	0,6941	0,3656

Uwagi: d – średnie szerokości połowy przedziałów ufności dla $\alpha = 0{,}05$, σ_b – średnie oszacowania odchylenia standardowego.

Źródło: badania własne.

Tabela 5.

Średnie szerokości połowy przedziałów ufności estymatora $X^{*b}_{(m_1,m_1+1)}$ dla $\alpha = 0{,}05$ oraz średnie oszacowania odchylenia standardowego wyznaczone z 1000 prób dla n parzystego

n	F(20,10) $M_{0,5}$ = 1,0349 σ = 0,8539		F(2,10) $M_{0,5}$ = 0,7435 σ = 1,6137		CHI(2) $M_{0,5}$ = 1,3863 σ = 2,0000		CHI(10) $M_{0,5}$ = 9,3418 σ = 4,4721		N(5,1) $M_{0,5}$ = 5 σ = 1,000		N(5,4) $M_{0,5}$ = 5 σ = 4,0000	
	d	σ_b	d	σ_b	d	σ_b	d	σ_b	d	σ_b	d	σ_b
10	0,6594	0,3108	1,1047	0,5438	1,7056	0,8179	4,3241	1,9970	0,9980	0,4567	3,9919	1,8268
20	0,4605	0,1963	0,7455	0,3249	1,2316	0,5382	3,2441	1,3788	0,7568	0,3169	3,0270	1,2676
30	0,3424	0,1536	0,5526	0,2514	0,9293	0,4220	2,4481	1,0962	0,5671	0,2531	2,2685	1,0123
40	0,3108	0,1261	0,4986	0,2033	0,8453	0,3459	2,2435	0,9154	0,5214	0,2132	2,0858	0,8528
50	0,2299	0,1132	0,3674	0,1825	0,6287	0,3116	1,6743	0,8226	0,3889	0,1908	1,5554	0,7631
60	0,2081	0,0984	0,3306	0,1572	0,5690	0,2703	1,5271	0,7224	0,3560	0,1685	1,4241	0,6739
70	0,2057	0,0946	0,3274	0,1514	0,5631	0,2602	1,5076	0,6934	0,3509	0,1614	1,4037	0,6456
80	0,1830	0,0875	0,2925	0,1413	0,5027	0,2419	1,3377	0,6359	0,3100	0,1466	1,2399	0,5863
90	0,1740	0,0810	0,2765	0,1292	0,4773	0,2229	1,2796	0,5959	0,2978	0,1386	1,1912	0,5546
100	0,1509	0,0753	0,2399	0,1204	0,4142	0,2073	1,1103	0,5518	0,2583	0,1282	1,0333	0,5126
110	0,1560	0,0757	0,2481	0,1211	0,4284	0,2087	1,1469	0,5550	0,2666	0,1287	1,0662	0,5149
120	0,1428	0,0701	0,2275	0,1116	0,3928	0,1928	1,0489	0,5161	0,2434	0,1201	0,9734	0,4805
130	0,1448	0,0686	0,2298	0,1094	0,3975	0,1890	1,0675	0,5053	0,2485	0,1175	0,9942	0,4701
140	0,1353	0,0645	0,2153	0,1030	0,3722	0,1779	0,9957	0,4745	0,2312	0,1101	0,9248	0,4406
150	0,1234	0,0628	0,1958	0,1000	0,3391	0,1730	0,9101	0,4626	0,2117	0,1075	0,8468	0,4300
160	0,1244	0,0606	0,1974	0,0964	0,3420	0,1669	0,9185	0,4468	0,2138	0,1039	0,8551	0,4157
170	0,1177	0,0589	0,1868	0,0938	0,3237	0,1623	0,8689	0,4344	0,2021	0,1010	0,8085	0,4040
180	0,1177	0,0558	0,1867	0,0888	0,3236	0,1538	0,8690	0,4118	0,2022	0,0958	0,8088	0,3830
190	0,1124	0,0551	0,1783	0,0877	0,3091	0,1519	0,8307	0,4070	0,1933	0,0947	0,7733	0,3786
200	0,1063	0,0537	0,1686	0,0853	0,2923	0,1478	0,7851	0,3960	0,1827	0,0921	0,7307	0,3683

Uwagi: d – średnie szerokości połowy przedziałów ufności dla α=0,05, σ_b – średnie oszacowania odchylenia standardowego.

Źródło: badania własne.

W przypadku estymatora $X^{*a}_{(m_1,m_1+1)}$ wzrostowi liczebności próby dla wszystkich badanych rozkładów i wszystkich wziętych pod uwagę liczebności towarzyszy zmniejszanie średniej szerokości przedziałów ufności. Jeśli chodzi o średnie odchylenie standardowe, to jedynie dla liczebności 110 zaobserwowano wzrost jego wartości wobec odchylenia dla $n = 100$. Dla estymatora $X^{*b}_{(m_1,m_1+1)}$ wzrost średniej długości przedziału wystąpił dla $n = 110$, 130 i 160 dla wszystkich wziętych pod uwagę rozkładów, zaś dla rozkładów *CHI*(10), *N*(5,1) i *N*(5,4) jeszcze dla $n = 180$. Wzrost średniego odchylenia standardowego zaobserwowano jedynie dla $n = 110$.

Relacja połowy średniej długości przedziału ufności dla $\alpha = 0{,}05$ do średniego odchylenia standardowego estymatora $X^{*a}_{(m_1,m_1+1)}$ jest równa 1,88, a dla estymatora $X^{*b}_{(m_1,m_1+1)}$ zaś 2,11. Może to wskazywać, że rozkład estymatora $X^{*a}_{(m_1,m_1+1)}$ jest bliższy rozkładowi normalnemu niż estymatora $X^{*b}_{(m_1,m_1+1)}$.

Pruska (2007) badała miedzy innymi wpływ zwiększania liczebności próby o jeden, dla prób 20 i 40 elementowych, na oszacowanie odchylenia standardowego estymatora bootstrapowego w postaci mediany z próby. Z badań tych wynikało, że zwiększenie liczebności (do liczebności nieparzystej) skutkowało dla większości badanych rozkładów zmniejszaniem odchylenia standardowego.

Oszacowania odchylenia standardowego bootstrapowych estymatorów mediany w postaci mediany z próby przedstawione w tabelach 2, 4 i 5, nie wskazują na tak jednoznaczne wyniki. Dla rozkładu *F*(20,10) odchylenia standardowe, dla prób o liczebnościach nieparzystych o jeden większych niż o liczebnościach parzystych, były w 10 na 20 przypadków większe, dla *F*(2,10), *CHI*(2) i *CHI*(10) w 11, dla *N*(5,1) i *N*(5,4) w 13. Połowy szerokości przedziałów ufności zaś, były we wszystkich przypadkach większe dla prób o nieparzystej (większej o jeden) liczebności, co może wynikać ze znacznie większej liczby realizacji estymatora $X^{*a}_{(m_1,m_1+1)}$ od estymatora $X^{*}_{(m)}$. Oparte na podobnych zasadach porównanie odchyleń standardowych estymatorów $X^{*}_{(m)}$ i $X^{*b}_{(m_1,m_1+1)}$ dało wyniki przeciwne. Oszacowane odchylenia standardowe estymatora $X^{*}_{(m)}$ były mniejsze w większości przypadków niż estymatora $X^{*b}_{(m_1,m_1+1)}$, podobnie jak połowy długości przedziałów ufności.

6. PODSUMOWANIE

W artykule pokazano, że dokładna metoda percentyli może być wykorzystana do szacowania mediany. Stosując estymator standardowy, dla prób o wybranych nieparzystych i parzystych liczebnościach, wyznaczono numery obserwacji stanowiące granice przedziałów ufności. Pokazano, że nie jest w tym przypadku potrzebne losowanie prób wtórnych.

Dla wziętych pod uwagę rozkładów i liczebności prób, estymator w postaci mediany z próby o parzystej liczbie obserwacji dawał prawidłowe oszacowania prze-

działów ufności, które zawierały faktyczną medianę rozkładu ze zliczeniowym poziomem ufności zbliżonym do założonego.

Estymator w postaci mediany z próby o parzystej liczbie obserwacji ma mniejszą wariancję od estymatora standardowego i dawał dokładniejsze oszacowania mediany (krótsze przedziały ufności). Zastosowanie jego wymaga jednak znacznie większego nakładu obliczeń.

Zwiększenie liczebności próby o jeden (z parzystej do nieparzystej liczebności) skutkowało wzrostem odchylenia standardowego i długości przydziałów ufności w połowie lub więcej przypadków dla estymatora w postaci mediany z próby, zaś zwykle zmniejszeniem tych wielkości w przypadku estymatora standardowego.

LITERATURA

Baszczyńska A., Pekasiewicz D., (2010), Selected MethodS of Interval Estimation of the Median. The Analysis of Accuracy of Estimation, *Acta Universitatis Lodziensis. Folia Oeconomica*, 235, 21–29.

Domański C., Pruska K., (2000), *Nieklasyczne metody statystyczne*, Polskie Wydawnictwo Ekonomiczne, Warszawa.

Efron B., (1979), Bootstrap Methods: Another Look at the Jackknife, *The Annals of Statistics*, 7 (1), 1–26.

Evans D. L., Leemis L. M., Drew J. H., (2006), The Distribution of Order Statistics for Discrete Random Variables With Applications of Bootstrapping, *Journal on Computing*, 18 (1), 19–30.

Fisher N. I., Hall P., (1991), Bootstrap Algorithms for Small Samples, *Journal of Statistical Planning and Inference*, 27, 157–169.

Hutson A. D., Ernst M. D., (2000), The Exact Bootstrap Mean and Variance of an *L*-estimator, *Journal of the Royal Statistical Society: Series B*, 62 (1), 89–94.

Kisielińska J., (2013), The Exact Bootstrap Method Shown on the Example of the Mean and Variance Estimation, *Computational Statistics*, 28 (3), 1061–1077.

Maritz J. S., Jarrett R. G., (1978), A Note on Estimating the Variance of the Sample Median, *Journal of the American Statistical Association*, 73 (361), 194–196.

Pekasiewicz D., (2015), *Statystyki pozycyjne w procedurach estymacji i ich zastosowania w badaniach ekonomicznych*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.

Pruska K., (2007), Estimation of Bias and Variance of Sample Median by Jackknife and Bootstrap Methods, *Acta Universitatis Lodziensis, Folia Oeconomica*, 206, 67–78.

Wilcox R. R., (2001), Fundamentals of Modern Statistical Methods, Springer, New York.

Zieliński R., (1995), Estimating Median and Other Quantiles in Nonparametric Models, *Applicationes Mathematicae*, 23 (3), 363–370.

ROZKŁADY WYBRANYCH BOOTSTRAPOWYCH ESTYMATORÓW MEDIANY ORAZ ZASTOSOWANIE DOKŁADNEJ METODY PERCENTYLI DO JEJ PRZEDZIAŁOWEGO SZACOWANIA

Streszczenie

W artykule przedstawiono wyniki zastosowania dokładnej metody percentyli do szacowania mediany. Dla próby o nieparzystej liczbie obserwacji zastosowano estymator w postaci mediany z próby (estymator standardowy), zaś dla próby o parzystej liczbie obserwacji estymator w postaci mediany z próby oraz

estymator standardowy. Przedstawiono formuły określające rozkład estymatorów standardowych oraz opisano w jaki sposób można wyznaczyć rozkład estymatora w postaci mediany z próby o parzystej liczbie obserwacji. Badania przeprowadzono dla wybranych rozkładów i wybranych liczebności próby. Oszacowano odchylenia standardowe wziętych pod uwagę estymatorów oraz przedziały ufności wyznaczone przy ich pomocy. W pracy pokazano, że użycie dokładnej metody percentyli dla estymatorów standardowych jest znacznie prostsze niż stosowanie metody klasycznej z losowaniem prób wtórnych.

Słowa kluczowe: dokładna metoda precentyli, estymacja mediany

DISTRIBUTION OF SELECTED BOOTSTRAP MEDIAN ESTIMATORS AND APPLICATION OF THE EXACT PERCENTILE METHOD FOR ITS INTERVAL ESTIMATING

Abstract

The article presents the results of the use of exact percentiles methods for estimating median. For a sample with an odd number of observations used estimator in the form of the sample median (standard estimator), and for a sample with an even number of observations, used estimator in the form of the sample median and standard estimator. It presents a formula defining the distribution of standard estimator and describes how you can determine the distribution of the estimator in the form of the sample median for sample with an even number of observations. The study was conducted for the selected distributions and selected sample size. It was estimated standard deviation and confidence intervals for selected estimators. The study shows, that the use of exact percentiles method for standard estimators is much simpler than using the classical method with the resampling.

Keywords: exact percentiles method, median estimating

DOI: 10.5604/01.3001.0014.1224

PRZEGLĄD STATYSTYCZNY
R. LXIII – ZESZYT 4 – 2016

PIOTR SULEWSKI[1]

MOC TESTÓW NIEZALEŻNOŚCI W TABLICY TRÓJDZIELCZEJ 2×2×2

1. WPROWADZENIE

Moc testów to prawdopodobieństwo odrzucenia hipotezy zerowej, gdy nie jest ona prawdziwa. Testy niezależności należą do najczęściej stosowanych narzędzi statystycznych, a dane do nich aranżuje się w postaci tablic wielodzielczych, a w szczególności tablic trójdzielczych (TT) 2×2×2. Wydawać by się mogło, że o metodach analizy tablic wielodzielczych powiedziano już wszystko. Jednak w dobie szeroko rozwiniętej informatyzacji i możliwości przeprowadzania symulacji komputerowych, na wiele zagadnień można spojrzeć „nowym okiem".

W (Cressie, Read, 1984) zaproponowano dla tablic dwudzielczych (TD) $w \times k$ rodzinę 6 statystyk, które mają asymptotyczny rozkład chi-kwadrat z $(w-1)(k-1)$ stopniami swobody. Do tzw. „statystyk chi-kwadrat" należą następujące statystyki: χ^2 Pearsona (Pearson, 1900), G^2 ilorazu wiarygodności (Sokal, Rohlf, 2012), *N* Neymana (Neyman, 1949), *KL* Kullbacka-Leiblera (Kullback, 1959), *FT* Freemana-Tukeya (Freeman, Tukey, 1950) oraz *CR* Cressiego-Reada (Cressie, Read, 1984). W pracy Sulewskiego (2014) pokazano, że „statystyki chi-kwadrat" są wystarczająco bliskie rozkładowi chi-kwadrat dla próbek liczących przynajmniej kilkaset elementów. Aby można było skorzystać z próbek mniej licznych wyznaczono wartości krytyczne na drodze symulacji komputerowych metodą Monte Carlo. Wyznaczanie wartości krytycznych na podstawie np. 100.000 wartości statystyki testowej – w dobie wydajnych komputerów z procesorami wielordzeniowymi – nie stanowi problemu.

W pracy Sulewski (2013) zaproponowano statystykę modułową będącą modyfikacją statystyki χ^2 Pearsona dla TD $w \times k$. Statystykę modułową porównano pod względem mocy testu ze „statystykami chi-kwadrat" dla TD 2×2 (Sulewski, 2016a) oraz dla wybranych rozmiarów TD większych niż 2×2 (Sulewski, 2016b).

Celem pracy jest przedstawienie teorii dotyczącej autorskiej statystyki modułowej mierzącej niezależność zmiennych dla TT 2×2×2 i zbadanie jej własności w odniesieniu do znanych „statystyk chi-kwadrat", opisanie procedury generowania zawartości tych tablic metodą słupkową, wprowadzenie nowej miary nieprawdziwości H_0.

Praca składa się z siedmiu punktów. W punkcie drugim zdefiniowano TT 2×2×2 oraz opisano metodę słupkową generacji zawartości tych tablic. W punkcie trzecim

[1] Akademia Pomorska w Słupsku, Instytut Matematyki, ul. Arciszewskiego 22, 76-200 Słupsk, Polska, e-mail: piotr.sulewski@apsl.edu.pl.

rozszerzono postać „statystyk chi-kwadrat" oraz statystyki modułowej na TT 2×2×2. W punkcie czwartym opisano sposób wyznaczania wartości krytycznych. W punkcie piątym zaproponowano miarę nieprawdziwości hipotezy H_0 zdefiniowaną dla różnych schematów wyznaczania prawdopodobieństw p_{ijt} $(i,j,t = 1,2)$, dla których H_0 jest niesłuszna. Wartości proponowanej miary siły związku porównano z wartościami współczynników: Graya-Williamsa (Gray, Williams, 1975), Marcotorchino (Marcotorchino, 1984), Lombardo (Lombardo, 2011). W punkcie szóstym testy niezależności wykorzystujące statystykę modułową oraz „statystyki chi-kwadrat" porównano ze względu na ich moc – czyli zdolność testu do odrzucenia hipotezy H_0 mówiącej o tym, że nie ma związku między cechami *X*, *Y*, *Z* w sytuacji, gdy w rzeczywistości jest ona fałszywa. Punkt ostatni poświęcono podsumowaniu.

2. TABLICA TRÓJDZIELCZA 2×2×2

Tablicę, która gromadzi wynik podziału próby według trzech cech *X*, *Y*, *Z* nazywamy tablicą trójdzielczą (TT). Można sobie ją wyobrazić jako kostkę z „*w*" wierszami, „*k*" kolumnami i „*p*" płaszczyznami. TT można także zilustrować rozkładając płaszczyzny *p* obok siebie. TT ułatwia odczytywanie zależności między cechami tak ilościowymi, jak i jakościowymi. Najprostszą jej postacią jest tablica 2×2×2 (tabela 1), która składa się z 8 liczebności n_{ijt} $(i,j,t = 1,2)$ rozkładu łącznego cech *X*, *Y*, *Z*.

Tabela 1.

Tablica trójdzielcza 2×2×2

Z	Z_1		Z_2		Razem
$X \setminus Y$	Y_1	Y_2	Y_1	Y_2	
X_1	n_{111}	n_{121}	n_{112}	n_{122}	$n_{1\bullet\bullet}$
X_2	n_{211}	n_{221}	n_{212}	n_{222}	$n_{2\bullet\bullet}$
Razem	$n_{\bullet 11}$	$n_{\bullet 21}$	$n_{\bullet 12}$	$n_{\bullet 22}$	n

Źródło: opracowanie własne.

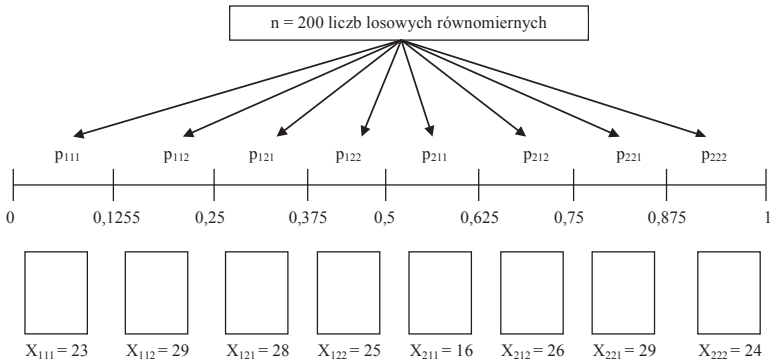
Liczebności oczekiwane dla TT 2×2×2 – przy założeniu niezależności badanych zmiennych – wyznacza się ze wzoru

$$e_{ijt} = \frac{n_{i\bullet\bullet} \cdot n_{\bullet j\bullet} \cdot n_{\bullet\bullet t}}{n^2} \quad (i, j, t = 1,2). \tag{1}$$

TT 2×2×2 można także przedstawić za pomocą prawdopodobieństw p_{ijt} $(i,j,t = 1,2)$ takich, że $\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{t=1}^{2} p_{ijt} = 1$. W odniesieniu do tabeli 1 wartości te wyznaczane są

ze wzoru $p_* = n_* / n$. Łatwo udowodnić, że sumy brzegowe liczebności oczekiwanych e_{ijt} są takie same jak sumy brzegowe liczebności n_{ijt}.

Generację zawartości TT 2×2×2 niezbędną do przeprowadzenia symulacji i wyznaczenia mocy testów przeprowadzono metodą słupkową (Sulewski, 2014). W tym celu przedział $\langle 0,1 \rangle$ podzielono na 2·2·2 podprzedziałów o szerokościach równych wartościom prawdopodobieństw p_{ijt} w taki sposób, że pierwszy podprzedział ma szerokość p_{111}, drugi – p_{112},…, ostatni – p_{222}. Wielkości p_{ijt} spełniają oczywiście warunek normalizacji $\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{t=1}^{2} p_{ijt} = 1$ i dobrano je w taki sposób, aby uzyskać żądaną wartość miary nieprawdziwości H_0. Każda z n wygenerowanych liczb losowych równomiernych „wpada" do jednego z podprzedziałów i tym samym zostaje o jedną zwiększona liczba obiektów w odpowiadającej temu podprzedziałowi komórce TT. Wielkości n_{ijt} spełniające równość $\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{t=1}^{2} n_{ijt} = n$ są liczebnością obiektów w poszczególnych komórkach TT. Rysunek 1 przedstawia schemat wypełnienia komórek TT 2×2×2 dla liczebności próby n = 200, gdy hipoteza zerowa jest słuszna, czyli p_{ijt} = 1/8 dla każdego i, j, t = 1,2. Tabela 2 prezentuje odpowiadającą temu schematowi TT 2×2×2.



Rysunek 1. Schematyczne ujęcie metody słupkowej
Źródło: opracowanie własne.

Tabela 2.

TT 2×2×2 uzyskana metodą słupkową

Z	Z_1		Z_2		Razem
$X \setminus Y$	Y_1	Y_2	Y_1	Y_2	
X_1	23	28	29	25	105
X_2	16	29	26	24	95
Razem	39	57	55	49	200

Źródło: opracowanie własne.

3. TESTY NIEZALEŻNOŚCI W TT 2×2×2

W analizie niezależności cech X, Y, Z można badać: niezależność pełną, niezależność brzegową, niezależność częściową, niezależność łączną oraz niezależność warunkową. W pracy niniejszej badania ograniczone zostały tylko do niezależności pełnej, którą można wyrazić zapisem (XYZ, XY, XZ, YZ).

Przy badaniu niezależności pełnej cech w tablicy 2×2×2 założono hipotezę zerową, że między cechami *X*, *Y*, *Z* związek nie występuje, tzn.

$$H_0 : p_{ijt} = p_{i\bullet\bullet} \cdot p_{\bullet j\bullet} \cdot p_{\bullet\bullet t} \tag{2}$$

dla każdego $i,j,t = 1,2$.

Do badania niezależności pełnej cech *X*, *Y*, *Z* w tablicy 2×2×2 skorzystano ze statystyki modułowej, która ma postać

$$|\chi| = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{t=1}^{2}\frac{|n_{ijt} - e_{ijt}|}{e_{ijt}} \tag{3}$$

oraz z rodziny statystyk, które są rozszerzeniem „statystyk chi-kwadrat" dla TD (Cressie, Read, 1984). Zalicza się do nich:

1) statystykę χ^2 Pearsona

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{t=1}^{2}\frac{(n_{ijt} - e_{ijt})^2}{e_{ijt}}, \tag{4}$$

2) statystykę G^2 ilorazu wiarygodności

$$G^2 = 2\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{t=1}^{2} n_{ijt} \ln\left(\frac{n_{ijt}}{e_{ijt}}\right), \tag{5}$$

3) statystykę *N* Neymana

$$N = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{t=1}^{2}\frac{(n_{ijt} - e_{ijt})^2}{n_{ijt}}, \tag{6}$$

4) statystykę *KL* Kullbacka-Leiblera

$$KL = 2\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{t=1}^{2} e_{ijt} \ln\left(\frac{e_{ijt}}{n_{ijt}}\right), \tag{7}$$

5) statystykę *FT* Freemana-Tukeya

$$FT = 4\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{t=1}^{2}\left(\sqrt{n_{ijt}} - \sqrt{e_{ijt}}\right)^2, \tag{8}$$

6) statystykę *CR* Cressiego-Reada

$$CR = \frac{9}{5}\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{t=1}^{2} n_{ijt}\left[\left(\frac{n_{ijt}}{e_{ijt}}\right)^{2/3} - 1\right]. \tag{9}$$

Powyższa rodzina statystyk dla TT 2×2×2 przy prawdziwości hipotezy H_0 ma asymptotyczny rozkład chi-kwadrat z 4 stopniami swobody i dlatego jest określana mianem „statystyk chi-kwadrat".

Charakterystykę statystyk „chi-kwadrat" dla tablic dwudzielczych, trójdzielczych i czterodzielczych wraz z implementacją komputerową można znaleźć w Sulewski (2014).

Liczebności n_{ijt} (i,j,t = 1,2) w TT 2×2×2 są zmiennymi losowymi. Godnym uwagi jest fakt, że zmienne losowe będące odwrotnościami lub ilorazami zmiennych losowych (patrz statystyki (6) i (7)) mogą nie posiadać wartości oczekiwanej i w konsekwencji także niektórych momentów wyższych rzędów. Rozkład chi-kwadrat – jak powszechnie wiadomo – posiada wszystkie momenty.

Ponadto pożądanym jest, aby poszczególne składniki statystyki testowej – gdy H_0 o niezależności cech X i Y jest słuszna – były tego samego znaku i jak najbliższe 0. Składniki statystyk (5), (7) i (9) są różnego znaku i wzajemnie się rekompensują i tylko dla prób liczących przynajmniej kilkaset elementów podlegają rozkładowi chi-kwadrat (Sulewski, 2014). Tabela 3 przedstawia wartości poszczególnych składników „statystyk chi-kwadrat" wyznaczonych symulacyjnie dla TT 2×2×2 i liczebności próby n = 200, gdy H_0 o niezależności cech X, Y, Z jest słuszna. Tabelę tę wygenerowano metodą słupkową opisaną powyżej dla p_{ijt} = 0,125 (i,j,t = 1,2). Wartości ujemne pogrubiono.

Tabela 3.

Wartości składników „statystyk chi-kwadrat"

Składnik	χ^2	G^2	N	KL	FT	CR
i = 1, j = 1, t = 1	0,667	**-7,297**	0,801	-8,758	0,730	**-4,123**
i = 1, j = 1, t = 2	0,260	-5,197	0,235	**-4,702**	0,247	-3,224
i = 1, j = 2, t = 1	0,010	-1,036	0,010	**-1,015**	0,010	-0,625
i = 1, j = 2, t = 2	0,041	-2,076	0,040	**-1,994**	0,041	-1,262
i = 2, j = 1, t = 1	2,573	-18,370	1,948	**-13,907**	2,228	12,111

Tabela 3. (cd.)

Składnik	χ^2	G^2	N	KL	FT	CR
$i = 2, j = 1, t = 2$	1,719	**-11,081**	2,338	-15,074	1,993	**-6,011**
$i = 2, j = 2, t = 1$	0,773	**-8,234**	0,932	-9,929	0,847	**-4,645**
$i = 2, j = 2, t = 2$	0,346	-6,328	0,310	**-5,674**	0,327	-3,938
Razem	6,389	-6,394	6,614	-6,468	6,422	-6,383

Źródło: opracowanie własne.

4. WYZNACZANIE WARTOŚCI KRYTYCZNYCH

W okresie ostatnich kilkudziesięciu lat zaproponowano różne testy niezależności dla tablic kontyngencji oraz określono warunki ich stosowalności szczególnie dla małych prób, które najczęściej są przedmiotem badań statystycznych. Warunki te głównie dotyczą bardzo popularnego i powszechnie stosowanego testu niezależności χ^2 Pearsona. W dobie powszechnie stosowanych i ciągle rozwijających się technologii komputerowych można za pomocą stosownego oprogramowania znieść te ograniczenia i drogą symulacyjną wyznaczyć wartości krytyczne.

Przy wyznaczaniu wartości krytycznych, gdy między cechami nie ma związku, zawartość TT 2×2×2 o liczebności próby n wygenerowano za pomocą metody słupkowej przyjmując $p_{ijt} = 0{,}125$ ($i,j,t = 1,2$).

Niech ϖ będzie jedną z rozpatrywanych siedmiu statystyk. Dla każdej TT 2×2×2 obliczono $r = 10^5$ razy wartości statystyki testowej ϖ, które następnie uporządkowano w kolejności rosnącej. Wartość krytyczną na poziomie istotności $\alpha = 0{,}1$ wyznaczono na podstawie dziewiątego decyla ze wzoru $cv_\alpha = \varpi_{(90000)}$. Tak duża liczba powtórzeń r przy wyznaczaniu wartości statystyki testowej zapewnia uzyskanie dokładnego wyniku.

5. MIARY SIŁY ZWIĄZKU DLA TT 2×2×2

Przeprowadzając badanie populacji generalnej istotna jest nie tylko zależność istniejąca między cechami, ale także jej siła. Teoria poświęcona miarom siły związku dla TT nie jest tak bogata jak dla TD, gdzie na szczególną uwagę zasługuje współczynnik τ Goodmana–Kruskala (Goodman, Kruskal, 1979). Numerycznymi rozszerzeniami tego współczynnika dla TT są: współczynnik Graya–Williamsa (Gray, Williams, 1975), współczynnik Marcotorchino (Marcotorchino, 1984) oraz współczynnik Lombardo (Lombardo, 2011). Informacje o innych mniej popularnych miarach można znaleźć w pracach Tuckera (1963), Harshmana (1970), Beha, Davy (1998) i Lombardo, Beha (2010).

Jeżeli dwie cechy (np. Y, Z) są cechami objaśniającymi lub predyktorami wzajemnie zależnymi (ang. *dependent predictor variables*), a trzecia cecha (np. X) jest

cechą objaśnianą lub odpowiedzi (ang. *response variable*), to współczynnik Graya–Williamsa ma postać:

$$\tau_{GW} = \left[\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{t=1}^{2} p_{\bullet jt}\left(p_{ijt}\cdot p_{\bullet jt}^{-1} - p_{i\bullet\bullet}\right)^2\right]\cdot\left[1-\sum_{i=1}^{2} p_{i\bullet\bullet}^2\right]^{-1}, \tag{10}$$

gdzie $p_{*} = n_{*} / n$. Współczynnik ten przyjmuje wartości z przedziału $\langle 0,1\rangle$. Jeżeli cechy Y i Z nie dostarczają informacji o cesze X, wtedy $\tau_{GW} = 0$. Wraz ze wzrostem informacji dostarczanej przez cechy Y i Z o cesze X wartość tego współczynnika zbliża się do jedności.

Jeżeli dwie cechy (np. Y, Z) są cechami objaśniającymi lub predyktorami wzajemnie niezależnymi (ang. *independent predictor variables*), a trzecia cecha (np. X) jest cechą objaśnianą lub odpowiedzi (*response variable*), to współczynnik Marcotorchino ma postać:

$$\tau_{M} = \left[\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{t=1}^{2} p_{\bullet j\bullet}\cdot p_{\bullet\bullet t}\left(p_{ijt}\cdot p_{\bullet\bullet t}^{-1}\cdot p_{\bullet j\bullet}^{-1} - p_{i\bullet\bullet}\right)^2\right]\cdot\left[1-\sum_{i=1}^{2} p_{i\bullet\bullet}^2\right]^{-1}, \tag{11}$$

gdzie $p_{*} = n_{*} / n$. Jeżeli $p_{ijt} = p_{i\bullet\bullet}p_{\bullet j\bullet}p_{\bullet\bullet t}$, to $\tau_{M} = 0$, natomiast gdy $p_{ijt} = p_{i\bullet\bullet}p_{\bullet\bullet t}$, to $\tau_{M} = 1$.

Jeżeli dwie cechy (np. X, Y) są cechami objaśnianymi lub odpowiedzi (ang. *independent response variables*) a trzecia cecha (np. Z) jest cechą objaśniającą lub predyktorem, to współczynnik Lombardo – zwany także współczynnikiem delta – ma postać

$$\tau_{L} = \left[\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{t=1}^{2} p_{\bullet\bullet t}\left(p_{ijt}\cdot p_{\bullet\bullet t}^{-1} - p_{i\bullet\bullet}\cdot p_{\bullet j\bullet}\right)^2\right]\cdot\left[1-\sum_{j=1}^{2} p_{i\bullet\bullet}^2\cdot\sum_{i=1}^{2} p_{\bullet j\bullet}^2\right]^{-1}, \tag{12}$$

gdzie $p_{*} = n_{*} / n$. Jeżeli $p_{ijt} = p_{i\bullet\bullet}p_{\bullet j\bullet}p_{\bullet\bullet t}$, to $\tau_{L} = 0$. Wartość maksymalna $\tau_{L} = 1$, gdy $p_{1jt} = 0$ dla pewnych $(j,t = 1,2)$ i $p_{2bc} = 0{,}5$ dla pewnych $(b,c = 1,2; b \neq j \wedge c \neq t)$.

Więcej informacji na temat trzech powyższych współczynników – w tym także ich implementację komputerową w języku VBA – można znaleźć w pracach Sulewski (2014, 2015).

Bazując na klasycznej definicji niezależności cech X, Y, Z proponuje się miarę nieprawdziwości H_0 w postaci:

$$mn = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{t=1}^{2}\left|p_{ijt} - p_{i\bullet\bullet}\cdot p_{\bullet j\bullet}\cdot p_{\bullet\bullet t}\right|. \tag{13}$$

Hipoteza zerowa H_0 – mówiąca o tym, że między cechami X, Y, Z w TT 2×2×2 nie ma związku – jest słuszna, gdy $p_{ijt} = p_{i\bullet\bullet}\cdot p_{\bullet j\bullet}\cdot p_{\bullet\bullet t}$ dla $i,j,t = 1,2$. Zatem miara (13) przyjmuje wartość 0, gdy hipoteza H_0 jest słuszna. Im większe wartości *mn*, tym większa jest możliwość fałszywości H_0.

Miarę *mn* jako naturalną – wynikającą z definicji niezależności cech – wykorzystano w procesie wyznaczania mocy testów.

6. MOC TESTU

W celu wyznaczenia mocy testu – czyli zdolności TT do odrzucenia hipotezy zerowej H_0 mówiącej o tym, że nie ma związku między cechami X, Y, Z – gdy w istocie związek jest, niezbędna staje się generacja zawartości TT. Uwzględniając narzuconą siłę związku między cechami uzyskaną dzięki generatorowi liczb losowych i wyrażoną za pomocą miary nieprawdziwości mn, do wypełnienia TT 2×2×2 skorzystano z metody „słupkowej" wykorzystującej prawdopodobieństwa p_{ijt} $(i,j,t = 1,2)$.

Znając wartość statystyki testowej ϖ dla różnych liczebności próby n wyznaczono moc testu za pomocą wzoru $m = q / r$, gdzie q określa liczbę tych spośród $r = 10^5$ wszystkich możliwych przypadków, kiedy to wartość statystyki testowej ϖ jest nie mniejsza od wartości krytycznej cv_α.

Do wyznaczenia mocy testu skorzystano ze schematów A, B, C, AB, AC, BC wyznaczania prawdopodobieństw p_{ijt} $(i,j,t = 1,2)$ (tabela 4).

Tabela 4.

Schematy prawdopodobieństw ukazujące związek między cechami X, Y, Z

Schemat A

Z	Z_1		Z_2	
$X \setminus Y$	Y_1	Y_2	Y_1	Y_2
X_1	$p_o - k \cdot \Delta p$	p_o	$p_o - k \cdot \Delta p$	p_o
X_2	p_o	$p_o + k \cdot \Delta p$	p_o	$p_o + k \cdot \Delta p$

Schemat B

Z	Z_1		Z_2	
$X \setminus Y$	Y_1	Y_2	Y_1	Y_2
X_1	$p_o - k \cdot \Delta p$	p_o	$p_o - k \cdot \Delta p$	p_o
X_2	$p_o + k \cdot \Delta p$	p_o	$p_o + k \cdot \Delta p$	p_o

Schemat C

Z	Z_1		Z_2	
$X \setminus Y$	Y_1	Y_2	Y_1	Y_2
X_1	$p_o - k \cdot \Delta p$	$p_o + k \cdot \Delta p$	$p_o - k \cdot \Delta p$	$p_o + k \cdot \Delta p$
X_2	$p_o + k \cdot \Delta p$	$p_o - k \cdot \Delta p$	$p_o + k \cdot \Delta p$	$p_o - k \cdot \Delta p$

Schemat AB

Z	Z_1		Z_2	
$X \setminus Y$	Y_1	Y_2	Y_1	Y_2
X_1	$p_o - k \cdot \Delta p$	p_o	$p_o - k \cdot \Delta p$	p_o
X_2	p_o	$p_o + k \cdot \Delta p$	$p_o + k \cdot \Delta p$	p_o

Schemat AC

Z	Z_1		Z_2	
$X \setminus Y$	Y_1	Y_2	Y_1	Y_2
X_1	$p_o - k \cdot \Delta p$	p_o	$p_o - k \cdot \Delta p$	$p_o + k \cdot \Delta p$
X_2	p_o	$p_o + k \cdot \Delta p$	$p_o + k \cdot \Delta p$	$p_o - k \cdot \Delta p$

Schemat BC

Z	Z_1		Z_2	
$X \setminus Y$	Y_1	Y_2	Y_1	Y_2
X_1	$p_o - k \cdot \Delta p$	p_o	$p_o - k \cdot \Delta p$	$p_o + k \cdot \Delta p$
X_2	$p_o + k \cdot \Delta p$	p_o	$p_o + k \cdot \Delta p$	$p_o - k \cdot \Delta p$

Gdzie $p_o = 0{,}125$, $\Delta p = 0{,}125 \times 10^{-3}$, $k = 0,\ldots,1000$.

Źródło: opracowanie własne.

Wartość minimalna miar siły związku (13) wynosi zero, dlatego w tabeli 5 przedstawiono tylko wartości maksymalne czterech miar siły związku dla różnych schematów prawdopodobieństw. Moc testu została wyznaczona dla wartości miary *mn* w punktach $u = 0{,}1,\ldots,9$. Miara ta dla wszystkich schematów prawdopodobieństw zmienia się ze stałym krokiem. W celu uzyskania takiej sytuacji należało odpowiednio dobrać wartości indeksu nieprawdziwości k (tabela 6).

Tabela 5.

Wartości maksymalne miar siły związku

Miara	A	B	C	AB	AC	BC
τ_{GW}	0,111	0,333	1	0,222	0,644	0,733
τ_M	0,111	0,333	1	0,378	0,716	0,733
τ_L	0,026	0,091	0,333	0,099	0,241	0,234
mn	0,25	0,5	1	0,44	0,67	0,75

Źródło: opracowanie własne.

Tabela 6.

Wartość indeksu nieprawdziwości k w punkcie u

u	A	B, C, BC	AB	AC
0	0	0	0	0
1	316	100	115	107
2	448	200	226	211
3	548	300	334	314
4	633	400	438	416

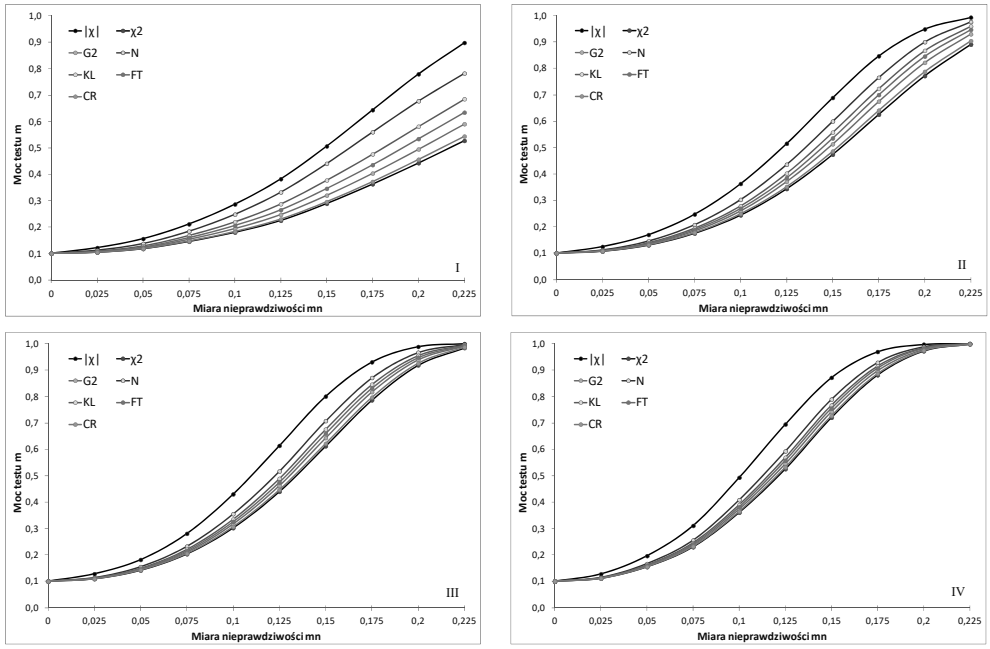
Tabela 6. (cd.)

u	A	B, C, BC	AB	AC
5	707	500	539	516
6	775	600	636	615
7	837	700	733	713
8	895	800	826	813
9	949	900	917	904
10	1000	1000	1000	1000

Źródło: opracowanie własne.

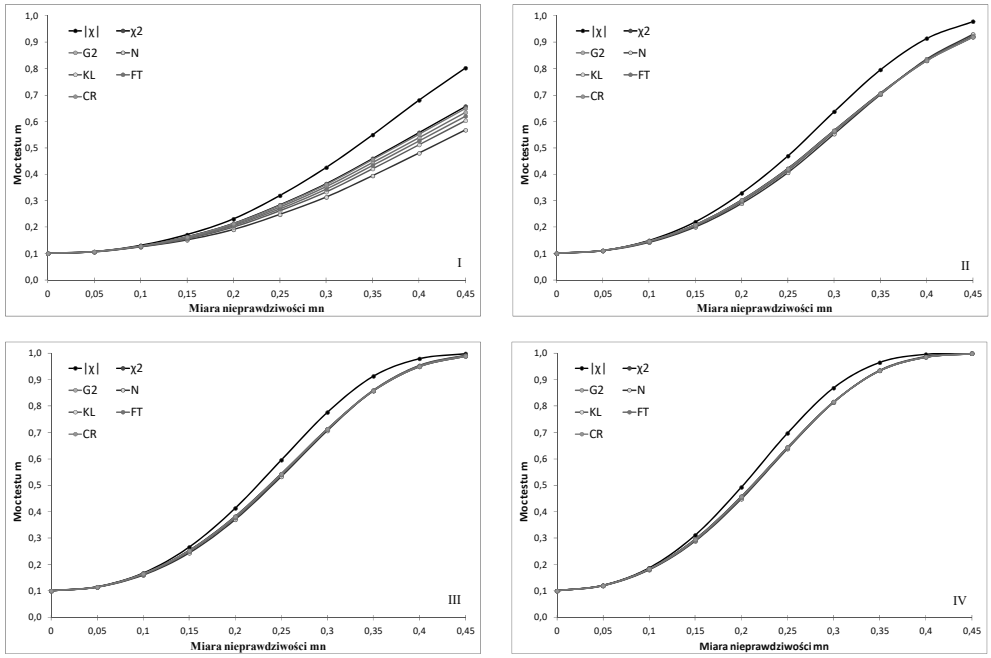
Aby można było porównać analizowane testy niezależności, liczebności komórek TT muszą być niezerowe, czyli $n_{ijt} \neq 0$ dla każdego $i,j,t = 1,2$, w sytuacji przeciwnej nie da się obliczyć wartości statystyk G^2, N, KL. W związku z powyższym moc testów niezależności zostanie obliczona dla $u = 0,1,\ldots,9$. Minimalną liczebność próby n dla danego schematu dobrano tak, aby $n_{ijt} \neq 0$ dla każdego $i,j,t = 1,2$. Maksymalną liczebność próby ustalono tak, aby uzyskać maksymalną moc testu.

Rysunki 2–7 przedstawiają zależność mocy testów niezależności od miary nieprawdziwości H_0 dla różnych schematów prawdopodobieństw, na poziomie istotności $\alpha = 0,1$ i liczebności próby n.



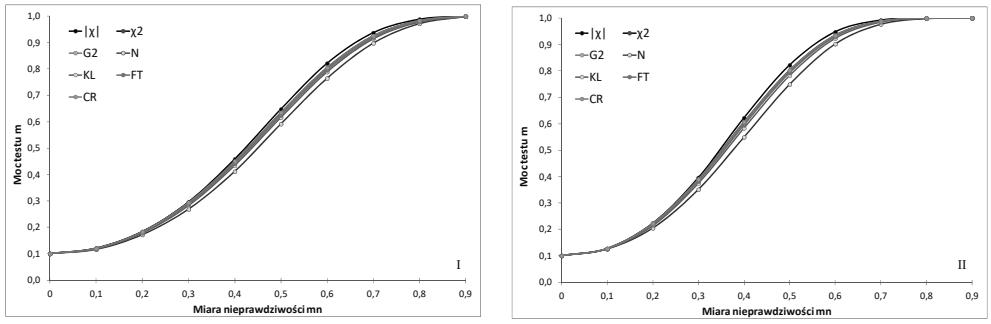
Rysunek 2. Moc testów niezależności dla schematu A prawdopodobieństw i liczebności próby $n = 40$ (I), $n = 60$ (II), $n = 80$ (III), $n = 100$ (IV)

Źródło: opracowanie własne.



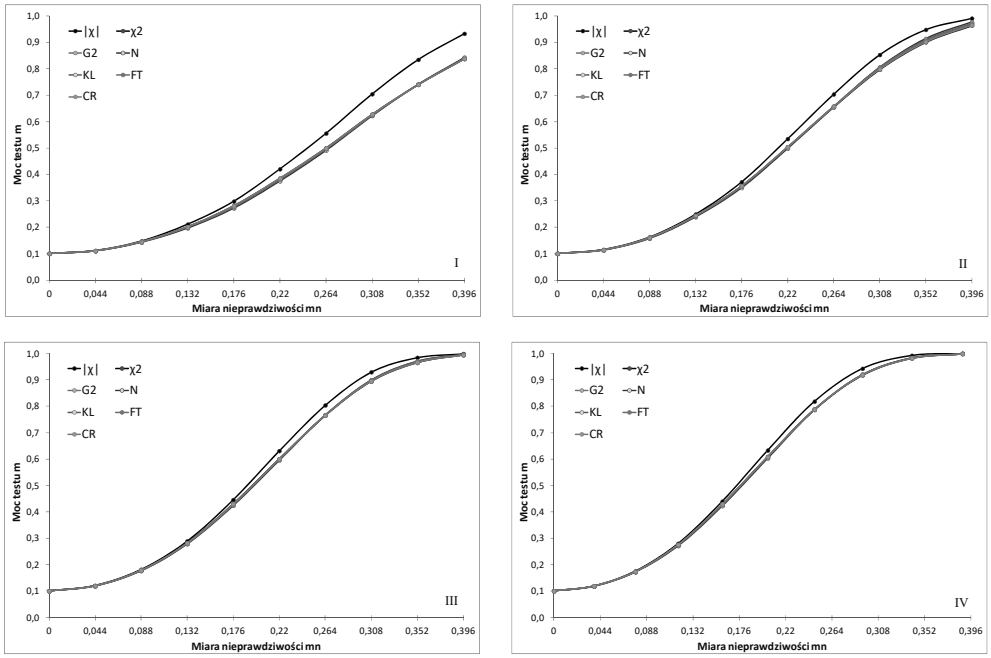
Rysunek 3. Moc testów niezależności dla schematu B prawdopodobieństw i liczebności próby $n = 40$ (I), $n = 60$ (II), $n = 80$ (III), $n = 100$ (IV)

Źródło: opracowanie własne.



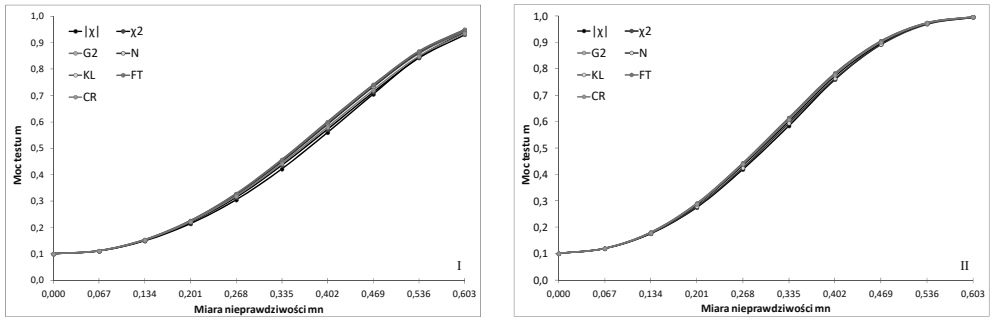
Rysunek 4. Moc testów niezależności dla schematu C prawdopodobieństw i liczebności próby $n = 30$ (I), $n = 40$ (II)

Źródło: opracowanie własne.



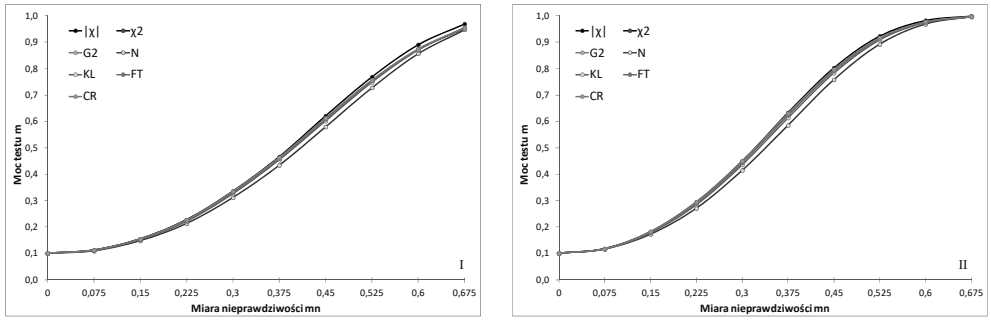
Rysunek 5. Moc testów niezależności dla schematu AB prawdopodobieństw i liczebności próby $n = 60$ (I), $n = 80$ (II), $n = 100$ (III), $n = 120$ (IV)

Źródło: opracowanie własne.



Rysunek 6. Moc testów niezależności dla schematu AC prawdopodobieństw i liczebności próby $n = 30$ (I), $n = 40$ (II)

Źródło: opracowanie własne.



Rysunek 7. Moc testów niezależności dla schematu BC prawdopodobieństw i liczebności próby $n = 30$ (I), $n = 40$ (II)

Źródło: opracowanie własne.

Dla statystyki modułowej $|\chi|$ oraz statystyki χ^2 Pearsona jako reprezentanta „statystyk chi-kwadrat", uzyskane wyniki porównano za pomocą testu o równości prawdopodobieństw stawiając hipotezę zerową, że moce testów są równe. Symbol „+" oznacza, że różnice między mocami testów są statystycznie nieistotne, symbol „-" – są statystycznie istotne (tabele 7–9).

Tabela 7.

Istotność statystyczna mocy testów $|\chi|$ i χ^2 dla schematów A i B

Schemat A

mn	$n = 100$	$n = 150$	$n = 200$	$n = 250$
0	+	+	+	+
0,025	+	+	+	+
0,05	-	-	-	-
0,075	-	-	-	-
0,1	-	-	-	-
0,125	-	-	-	-
0,15	-	-	-	-
0,175	-	-	-	-
0,2	-	-	-	-
0,225	+	-	+	+

Schemat B

mn	$n = 40$	$n = 60$	$n = 80$	$n = 100$
0	+	+	+	+
0,05	+	+	+	+
0,1	+	+	+	+
0,15	+	+	+	+
0,2	+	+	+	+
0,25	-	-	-	-
0,3	-	-	-	-
0,35	-	-	-	-
0,4	-	-	-	-
0,45	-	-	-	+

Źródło: opracowanie własne.

Tabela 8.

Istotność statystyczna mocy testów $|\chi|$ i χ^2 dla schematów C i AB

Schemat C

mn	*n* = 30	*n* = 40
0	+	+
0,1	+	+
0,2	+	+
0,3	+	+
0,4	+	+
0,5	+	+
0,6	+	+
0,7	+	+
0,8	+	+
0,9	+	+

Schemat AB

mn	*n* = 60	*n* = 80	*n* = 100	*n* = 120
0	+	+	+	+
0,044	+	+	+	+
0,088	+	+	+	+
0,132	+	+	+	+
0,176	+	+	+	+
0,22	+	+	+	+
0,264	-	-	-	-
0,308	-	-	-	-
0,352	-	-	-	-
0,396	+	+	+	+

Źródło: opracowanie własne.

Tabela 9.

Istotność statystyczna mocy testów $|\chi|$ i χ^2 dla schematów AC i BC

Schemat AC

mn	*n* = 30	n = 40
0	+	+
0,067	+	+
0,134	+	+
0,201	+	+
0,268	+	+
0,335	+	+
0,402	+	+
0,469	+	+
0,536	+	+
0,603	+	+

Schemat BC

mn	*n* = 30	*n* = 40
0	+	+
0,075	+	+
0,15	+	+
0,225	+	+
0,3	+	+
0,375	+	+
0,45	+	+
0,525	+	+
0,6	+	+
0,675	+	+

Źródło: opracowanie własne.

Z rysunków 2–7 wynika, że testy niezależności wykorzystujące „statystyki chi-kwadrat" we wszystkich schematach prawdopodobieństw charakteryzują się podobną mocą dla danej liczebności próby *n* i miary *mn*. Wyjątkiem jest tylko schemat A dla którego miara *mn* przyjmuje najmniejsze wartości. Test wykorzystujący statystykę modułową ma podobną moc jak pozostałe testy niezależności dla schematów C, AC, BC. Dla schematu A test wykorzystujący statystykę modułową ma większą moc niż testy związane ze „statystykami chi-kwadrat" i ta przewaga jest statystycznie istotna dla $mn > 0{,}025$. Podobna sytuacja jest także dla schematu B, gdy $mn > 0{,}02$ oraz dla schematu AB, gdy $mn > 0{,}022$.

7. PODSUMOWANIE

Kwadrat użyty w liczniku statystyki χ^2 powoduje, że duże rozbieżności między liczebnością empiryczną a teoretyczną są jeszcze większe, a małe rozbieżności jeszcze mniejsze. Innym celem zastosowania kwadratu było uniknięcie wzajemnego niwelowania się rozbieżności. Do tego jednak celu zamiast kwadratu odchyleń można użyć także ich wartości bezwzględnej, stąd pomysł na statystykę modułową.

Artykuł zwraca uwagę na fakt, że testy niezależności wykorzystujące „statystyki chi-kwadrat" we wszystkich schematach prawdopodobieństw charakteryzują się podobną mocą dla danej liczebności próby *n* i miary *mn*. Nieznacznym wyjątkiem od tej reguły jest schemat A, w którym miara nieprawdziwości *mn* przyjmuje najmniejsze wartości. Statystyka modułowa charakteryzuje się podobną mocą jak „statystyki chi-kwadrat" dla schematów C, AC, BC oraz większą mocą dla schematów A, B, AB. Przewaga na korzyść nowej statystyki szczególnie jest widoczna dla mniejszych liczebności próby w ramach danego schematu.

Badając niezależność trzech cech w pewnych schematach prawdopodobieństwa dla statystyki modułowej uzyskano wyższą moc wykrywania zależności, rozumianej w sensie zaproponowanej miary *mn*. Na korzyść proponowanej statystyki przemawia także fakt, że można z niej skorzystać w sytuacji, gdy komórki TT są puste. Statystyki G^2 ilorazu wiarygodności, *N* Neymana i *KL* Kullbacka-Leiblera nie spełniają tego założenia.

Zatem można wnioskować, że statystyka modułowa jest nie mniej skuteczna niż „statystyki chi-kwadrat" i może stanowić alternatywę, zwłaszcza uwzględniając jej inne zalety.

LITERATURA

Beh E. J., Davy P. J., (1998), Partitioning Pearson's Chi-squared Statistic for a Completely Ordered Three-way Contingency Table, *The Australian and New Zealand Journal of Statistics*, 40, 465–477.

Cressie N., Read T., (1984), Multinomial Goodness-of-Fit Tests, *Journal of the Royal Statistical Society*. Series B (Methodological), 46 (3), 440–464.

Freeman M. F., Tukey J. W., (1950), Transformations Related to the Angular and the Square Root, *Annals of Mathematical Statistics*, 21, 607–611.

Goodman L., Kruskal W., (1954), Measures of Association for Cross Classifications, *Journal of the American Statistical Association*, 49, 732–764.
Gray L. N., Williams J. S., (1975), Goodman and Kruskal's Tau b: Multiple and Partial Analogs, in: *Proceedings of the Social Statistics Section, American Statistical Association*.
Harshman R. A., (1970), Foundations of the PARAFAC Procedure: Models and Conditions for an Explanatory Multi-modal Factor Analysis, *UCLA Working Papers in Phonetics*, 16, 1–84.
Kullback S., (1959), *Information Theory and Statistics*, Wiley, New York.
Lombardo R., (2011), Three-way Association Measure Decompositions: The Delta index, *Journal of Statistical Planning and Inference*, 141, 1789–1799.
Lombardo R., Beh E. J., (2010), Simple and Multiple Correspondence Analysis for Ordinal-scale Variables Using Orthogonal Polynomials, *Journal of Applied Statistics*, 37 (12), 2101–2116.
Marcotorchino F., (1984), *Utilisation des Comparaisons par Paires en Statistique des Contingences*, Partie (I), *Etude IBM F069*, France.
Neyman J., (1949), Contribution to the Theory of the χ^2 Test, *Proceedings of the [First] Berkeley Symposium on Mathematical Statistics and Probability*, (Univ. of Calif. Press), 239–273.
Pearson K., (1900), On the Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables is such that it can be Reasonably Supposed to Have Arisen from Random Sampling, *Philosophy Magazine Series*, 5 (50), 157–172.
Sokal R. R., Rohlf F. J., (2012), *Biometry: the principles and practice of statistics in biological research*, Freeman, New York.
Sulewski P., (2013), Modyfikacja testu niezależności, *Wiadomości Statystyczne*, GUS, 10, 1–19.
Sulewski P., (2014), *Statystyczne badanie współzależności cech typu dyskretne kategorie*, Akademia Pomorska, Słupsk.
Sulewski P., (2015), Miary związku między cechami w tablicy trójdzielczej, *Wiadomości Statystyczne*, GUS, 1, 13–27.
Sulewski P., (2016a), Moc testów niezależności w tablicy dwudzielczej, *Wiadomości Statystyczne*, GUS, 8, 1–17.
Sulewski P., (2016b), Moc testów niezależności w tablicy dwudzielczej większej niż 2×2, *Przegląd Statystyczny*, 63 (2), 191–209.
Tucker L. R., (1963), Implications of Factor Analysis of Three-way Matrices for Measurement of Change, w: Harris C. W. (red.), *In Problems in Measuring Change*, 122–137.

MOC TESTÓW NIEZALEŻNOŚCI W TABLICY TRÓJDZIELCZEJ 2×2×2

Streszczenie

Pierwszym celem pracy jest przedstawienie teorii dotyczącej autorskiej statystyki modułowej mierzącej niezależność zmiennych dla tablic trójdzielczych 2×2×2 i zbadanie jej własności w odniesieniu do znanych „statystyk chi-kwadrat". Drugim celem jest opisanie procedury generowania zawartości tych tablic metodą słupkową. Trzecim celem jest zaproponowanie miary nieprawdziwości hipotezy zerowej, a także porównanie jakości testów niezależności za pomocą ich mocy. Wartości krytyczne dla testów niezależności wyznaczono symulacyjnie metodami Monte Carlo.

Słowa kluczowe: statystyka modułowa, tablica trójdzielcza, test niezależności, metoda słupkowa, Monte Carlo

POWER ANALYSIS OF INDEPENDENCE TESTING FOR THREE-WAY CONTINGENCY TABLE 2×2×2

A b s t r a c t

The first aim of this paper is to present the theory of the proposal of the author in the form of modular statistics for three-way contingency table 2×2×2 and examine its properties in relation to known "chi-squared statistics". The second aim is to describe the procedure of generating the content of these tables using the bar method. The third aim is to propose the measure of untruthfulness of null hypothesis as well as to compare the quality of independence tests using their power. Critical values for all analyzed statistics were determined by simulation methods of Monte Carlo.

Keywords: three-way contingency tables, modular statistics, independence test, bar method, Monte Carlo method

DOI: 10.5604/01.3001.0014.1225

PRZEGLĄD STATYSTYCZNY
R. LXIII – ZESZYT 4 – 2016

SERGIUSZ HERMAN[1]

ANALIZA PORÓWNAWCZA WYBRANYCH METOD SZACOWANIA BŁĘDU PREDYKCJI KLASYFIKATORA

1. WSTĘP

Proces konstrukcji klasyfikatora obejmuje trzy zasadnicze etapy: wybór cech opisujących klasyfikowane obiekty (zmiennych predykcyjnych), wybór metody/modelu oraz finalną ocenę jego jakości. Jakość klasyfikatora utożsamiana jest z jego umiejętnością przewidywania (prognozowania) przynależności do rozważanych populacji obiektów, dla których przynależność ta nie jest znana. Miarą tak zdefiniowanej jakości może być błąd predykcji klasyfikatora. Badania poświęcone klasyfikacji obiektów koncentrują swoją uwagę na wykorzystywaniu coraz bardziej zaawansowanych metod klasyfikacji, przyjmując powszechnie stosowane metody szacowania błędu predykcji. Jednocześnie podkreśla się, iż niezależnie od stopnia zaawansowania metody klasyfikacji, jakość decyzji podjętych na podstawie skonstruowanego za jej pomocą modelu uzależniona jest od tego, jak wiarygodnie zostanie oszacowana jego zdolność predykcyjna (Isaksson i inni, 2008).

W literaturze wskazuje się, iż ocena zdolności predykcyjnej klasyfikatora powinna być dokonana na podstawie dużej, niezależnej próby obiektów, które nie zostały uwzględnione przy konstrukcji modelu. Często jednak, szczególnie w przypadku badań dotyczących polskiego rynku kapitałowego, trudno taką próbę uzyskać. Rozwiązaniem tej sytuacji jest wówczas wykorzystanie jednej z wielu, zaproponowanych w literaturze, metod szacowania błędu predykcji klasyfikatora. W zagranicznych publikacjach można spotkać opracowania, których celem jest ich empiryczna analiza porównawcza. I tak na przykład Wehberg, Schumacher (2004) wykorzystali w tym celu metodę resubstytucji, metodę walidacji krzyżowej oraz metody wielokrotnego repróbkowania. Molinaro i inni (2005) rozszerzyli wspomnianą listę porównywanych metod o prostą metodę podziału, Braga-Neto, Dougherty (2004) oraz Kim (2009) uwzględnili z kolei w swoich badaniach powtarzaną walidację krzyżową oraz powtarzaną prostą metodę podziału. Analizy te zostały przeprowadzone na podstawie zmiennych opisujących wyniki eksperymentów genetycznych oraz danych uzyskanych w wyniku symulacji. W polskiej literaturze brakuje podobnego opracowania o charakterze empirycznym,

[1] Uniwersytet Ekonomiczny w Poznaniu, Wydział Informatyki i Gospodarki Elektronicznej, Katedra Ekonometrii, al. Niepodległości 10, 61-875 Poznań, Polska, e-mail: sergiusz.herman@ue.poznan.pl.

poświęconego metodom szacowania błędu predykcji. Przegląd metod szacowania błędu predykcji klasyfikatora można znaleźć np. w pracach Gatnara (2001, 2008).

Celem artykułu jest dokonanie przeglądu oraz empirycznej analizy porównawczej wybranych metod szacowania błędu predykcji klasyfikatora, skonstruowanego z wykorzystaniem liniowej analizy dyskryminacyjnej. W analizie porównawczej estymatorów wykorzystano trzy miary tj. odchylenie standardowe, obciążenie oraz błąd średniokwadratowy. Zbadano, czy wyniki analizy uzależnione są od wielkości próby oraz metody wyboru zmiennych do modelu. W tym celu badanie przeprowadzono dla prób o różnej liczebności oraz wykorzystano trzy statystyczne metody wyboru zmiennych do modelu.

Wyniki przeprowadzonej analizy pozwolą wskazać metodę szacowania błędu predykcji klasyfikatora, która posiada najbardziej pożądane własności, na przykładzie problemu prognozowania upadłości spółek akcyjnych w Polsce.

2. BŁĄD PREDYKCJI KLASYFIKATORA I WYBRANE METODY JEGO SZACOWANIA

Na podstawie notacji zaproponowanej m.in. przez Efrona, Tibishiraniego (1997) zagadnienie konstrukcji klasyfikatora oraz szacowania błędu predykcji można przedstawić w sposób formalny. Pierwszym krokiem badania jest zgromadzenie danych dla obiektów tworzących próbę uczącą x. Próba ta składa się z n obiektów, z których każdy opisany jest wektorem $x_i = (t_i, y_i)$, gdzie t_i to wektor zmiennych (cech) opisujących i-ty obiekt, natomiast y_i jest zmienną określającą przynależność tego obiektu do populacji. Wykorzystując tak zdefiniowany zbiór uczący x konstruowany jest klasyfikator (model) $r_{\boldsymbol{x}}$, który w oparciu o wartości zawarte w wektorze cech t_i dla i-tego obiektu, pozwala określić przynależność tego obiektu do badanych populacji, oznaczanej przez $r_{\boldsymbol{x}}(t_i)$. W przypadku, gdy problem dotyczy klasyfikacji dychotomicznej (zmienne y_i oraz $r_{\boldsymbol{x}}(t_i)$ przyjmują wartość 0 lub 1) można w następujący sposób zdefiniować miarę rozbieżności między rzeczywistą a prognozowaną przynależnością i-tego obiektu:

$$Q[y_i, r_x(t_i)] = \begin{cases} 0 \; jeżeli \; r_x(t_i) = y_i, \\ 1 \; jeżeli \; r_x(t_i) \neq y_i. \end{cases} \tag{1}$$

Skonstruowany klasyfikator wykorzystywany jest do prognozowania przynależności obiektów, tworzących próbę testową. Oznaczając taki obiekt za pomocą $x_0 = (t_0, y_0)$ przedstawioną właśnie miarę rozbieżności (wzór 1) można dla niego zapisać w skrócie $Q[y_0, r_{\boldsymbol{x}}(t_0)]$. Porównuje ona rzeczywistą przynależność badanego obiektu y_0, z przynależnością ustaloną za pomocą klasyfikatora r_x skonstruowanego w oparciu o zbiór uczący x, przy wykorzystaniu wektora cech t_0 opisującego klasyfikowany obiekt.

Przyjmując, iż obserwacje $x_i = (t_i, y_i)$ z próby uczącej są próbą losową z pewnego rozkładu F oraz że obiekt z próby testowej $x_0 = (t_0, y_0)$ także charakteryzuje się tym rozkładem, można zdefiniować tzw. prawdziwy błąd predykcji (ang. *true error rate*) klasyfikatora za pomocą następującej wartości oczekiwanej:

$$Err = E_{0F}Q[y_0, r_x(t_0)]. \tag{2}$$

Klasyfikator został skonstruowany w oparciu o daną próbę uczącą *x*, stąd też błąd *Err* jest także określany mianem warunkowego błędu predykcji (ang. *conditional error rate*). Oszacowanie prawdziwego błędu predykcji wymaga posiadania dużej, niezależnej próby testowej. W przypadku, gdy badacz nie dysponuje nią, jego zadaniem jest oszacowanie prawdziwego błędu predykcji w oparciu o posiadaną próbę uczącą *x* (Efron, Tibshirani, 1997).

Najprostszym sposobem oszacowania prawdziwego błędu predykcji jest wykorzystanie wszystkich dostępnych obserwacji, zarówno w celu zbudowania, jak i oceny konstruowanego klasyfikatora. Oszacowaniem warunkowego błędu predykcji jest wówczas tzw. estymator resubstytucji (ang. *resubstitution estimator*):

$$\overline{err} = \frac{1}{n}\sum_{i=1}^{n} Q[y_i, r_x(t_i)]. \tag{3}$$

Zgodnie z tym zapisem, klasyfikator zostaje skonstruowany w oparciu o cały zbiór uczący *x*. Następnie dla każdego obiektu *i* zostaje porównana jego rzeczywista przynależność y_i z tą, prognozowaną za pomocą klasyfikatora – $r_x(t_i)$. Na tej podstawie zostaje obliczona średnia przedstawiona we wzorze (3). Uzyskany za pomocą przedstawionej metody estymator błędu predykcji jest nadmiernie „optymistyczny", to znaczy nie doszacowuje wartości ryzyka warunkowego (McLachlan, 1992, s. 339–340).

Drugą, bardzo często stosowaną w badaniach metodą szacowania warunkowego błędu predykcji klasyfikatora, jest tzw. prosta metoda podziału (ang. *split sample/ holdout metod*). Polega ona na jednokrotnym podziale dostępnych danych, zgodnie z wcześniej ustaloną proporcją *p*, na próbę uczącą i testową. Klasyfikator jest konstruowany w oparciu o pierwszą z nich. Następnie jest on wykorzystywany do prognozowania przynależności obiektów z próby testowej. Błąd predykcji wyrażany jest jako udział błędnie zaklasyfikowanych obiektów z próby testowej $x_0 = (t_0, y_0)$ w ogólnej liczebności tej próby. Zaletą tej metody jest jej prostota i fakt, że nie pociąga za sobą obszernych obliczeń, ponieważ klasyfikator jest konstruowany tylko raz. Jej wadą natomiast jest fakt, iż każdy obiekt przyporządkowany zostaje tylko do jednej z analizowanych prób. Ma to istotny wpływ na uzyskane wyniki, szczególnie w przypadku występowania w zgromadzonych danych obserwacji odstających. Z prostej metody podziału należy korzystać tylko wtedy, gdy dysponuje się dostatecznie szerokim zbiorem danych, które pozwolą na wyodrębnienie odpowiednio licznych, niezależnych zbiorów: uczącego i testowego (Ripley, 1996, s. 67).

W wielu dziedzinach nauki badania empiryczne muszą być przeprowadzane na podstawie zbiorów danych o ograniczonej, niewielkiej liczbie obserwacji. W tej sytuacji przedstawiona prosta metoda podziału, ze względu na wspomniane wady, nie gwarantuje wiarygodnych wyników. Z tego też powodu opracowano szereg metod

szacowania błędu predykcji, z których każda bazuje na odpowiednim, wielokrotnym podziale dostępnego zbioru danych na próbę uczącą oraz testową.

Pierwszą z takich metod jest walidacja krzyżowa (Lachenbruch, Mickey, 1968; Geisser, 1975). W metodzie tej klasyfikator konstruowany jest *n*-krotnie. Za każdym razem z dostępnej próby usuwany jest jeden obiekt, który wykorzystywany jest jako jednoelementowy zbiór testowy, reszta obiektów służy natomiast do uczenia modelu. Tym samym formułę wykorzystywaną przy szacowaniu warunkowego błędu predykcji można zapisać następująco:

$$\widehat{Err}^{cv1} = \frac{1}{n}\sum_{i=1}^{n} Q\left[y_i, r_{x_{(i)}}(t_i)\right], \tag{4}$$

gdzie $x_{(i)}$ oznacza próbę uczącą po usunięciu z niej *i*-tego obiektu. Opisana metoda szacowania błędu predykcji określana jest mianem walidacji krzyżowej typu „pozostaw jedną poza" (ang. *leave-one-out cross-validation*).

Innym wariantem tej metody jest *k*-krotna walidacja krzyżowa. W tym przypadku dostępna próba zostaje podzielona na *k* części. Następnie *k*-krotnie klasyfikator jest konstruowany na podstawie $k - 1$ części, oraz testowany na tej, nieuwzględnionej w uczeniu. Oszacowaniem błędu predykcji jest średnia z uzyskanych w ten sposób *k* wyników pośrednich. Zaletą walidacji krzyżowej jest fakt, iż każda z obserwacji zostaje uwzględniona zarówno przy szacowaniu modelu, jak i przy jego testowaniu. Wadą metody jest większy (w porównaniu z metodami wcześniej opisanymi) koszt obliczeniowy.

Losowy podział próby na *k* części w przypadku walidacji krzyżowej powoduje, iż otrzymany za pomocą tej metody estymator charakteryzuje wysoka zmienność – określana w literaturze mianem wariancji wewnętrznej (ang. *internal variance*) (Braga-Neto, Dougherty, 2004). Jednym ze sposobów jej redukcji jest wielokrotne powtórzenie całego procesu (tj. dzielenia dostępnej próby, konstruowania klasyfikatora oraz szacowania błędu predykcji). Metoda ta określana jest mianem powtarzanej *k*-krotnej walidacji krzyżowej (ang. *repeated cross-validation*)

Kolejną metodą szacowania błędu predykcji jest powtarzana prosta metoda podziału (ang. *repeated hold-out*) (Kim, 2009). Polega ona na losowym, *k*-krotnym podziale badanej próby (składającej się z *n* obiektów) na dwa podzbiory: uczący i testowy zgodnie z ustaloną proporcją *p*. Dla każdego powtórzenia $n \cdot p$ obserwacji jest przyporządkowywanych do zbioru testowego, pozostałe $n \cdot (1 - p)$ obiektów traktowane jest jako zbiór testowy. Błąd predykcji szacowany jest jako średnia z wszystkich *k* iteracji – losowań. Przewagą tej metody nad przedstawioną wcześniej walidacją krzyżową jest fakt, iż w tym przypadku możliwe jest uzyskanie większej liczby, różnych w stosunku do siebie, podziałów pierwotnego zbioru obserwacji.

Ostatnią grupą metod szacowania błędu predykcji, które zostały uwzględnione w badaniu są metody wielokrotnego repróbkowania (ang. *bootsrapping*). Ta grupa metod bazuje na generowaniu *B* prób typu bootstrap x^{*1}, x^{*2}, x^{*3}, …, x^{*B} w taki

sposób, że każda z nich powstaje poprzez *n*-krotne losowanie proste ze zwracaniem obiektów z dostępnej próby *n* obiektów $\{x_1, x_2, \ldots, x_n\}$. Próby te wykorzystywane są następnie jako próby uczące. Obiekty niewylosowane w kolejnych próbach stanowią próbę testową. Miarę rozbieżności między rzeczywistą a prognozowaną przynależnością *i*-tego obiektu, dla uczącej próby *b*, można wyrazić w następujący sposób:

$$Q_i^b = Q\left[y_i, r_{x^{*b}}(t_i)\right]. \tag{5}$$

Niech N_i^b odpowiada liczbie przypadków, w których obiekt x_i został wylosowany w *b*-tej próbie bootstrap. Dla każdego obiektu *i* oraz każdej próby bootstrap *b* określona zostanie następująca zmienna dychotomiczna:

$$I_i^b = \begin{cases} 1 \; jeżeli \; N_i^b = 0, \\ 0 \; jeżeli \; N_i^b > 0. \end{cases} \tag{6}$$

Biorąc pod uwagę powyższe oznaczenia, estymator błędu predykcji typu „pozostaw jedną poza" (ang. *leave-one-out bootstrapp*) można przedstawić następująco (Efron, 1983):

$$\widehat{Err}^{(1)} = \frac{\sum_i \sum_b I_i^b \, Q_i^b}{\sum_i \sum_b I_i^b}. \tag{7}$$

Badania pokazały, iż otrzymany w ten sposób estymator warunkowego błędu predykcji przeszacowuje jego wartość. Dzieje się tak, ponieważ każda z wylosowanych podprób uczących zawiera w przybliżeniu tylko 0,632*n* unikatowych (niepowtarzających się) obserwacji. Z tego też powodu Efron (1983) zaproponował jego modyfikację. W celu uniknięcia problemu przeszacowania, estymator opisany wzorem (7) został skorygowany o przedstawiony wcześniej, niedoszacowany estymator resubstytucji (wzór 3) w następujący sposób:

$$\widehat{Err}^{(0,632)} = 0{,}368\overline{err} + 0{,}632\widehat{Err}^{(1)}. \tag{8}$$

Jednak badania empiryczne przeprowadzone przez Efrona, Tibshiraniego (1997) wskazały na znaczącą wadę tego estymatora. Okazało się, że posiada on tendencję do nadmiernego, ujemnego obciążenia, zwłaszcza w przypadku klasyfikatorów mających tendencję do nadmiernego przeuczenia – dopasowania do obiektów, na podstawie których jest on konstruowany. W takich sytuacjach estymator resubstytucji przyjmuje wartości bliskie zeru, a tym samym wartość błędu predykcji (wzór 8) jest zbyt optymistyczna. Efron oraz Tibshirani zaproponowali, by zwiększyć wagę przypisaną estymatorowi $\widehat{Err}^{(1)}$ wtedy, kiedy skala nadmiernego dopasowania (przeuczenia) mierzona za pomocą różnicy między wartościami $\widehat{Err}^{(1)}$ a $\overline{err}$ wzrasta. Przedstawiono estymator następującej postaci (Efron, Tibshirani, 1997):

$$\widehat{Err}^{(0,632+)} = (1 - \hat{w}) \cdot \overline{err} + \hat{w} \cdot \widehat{Err}^{(1)}, \tag{9}$$

gdzie:
$\overline{err}$ – estymator opisany wzorem (3)

$$\hat{w} = \frac{0{,}632}{1 - 0{,}368\hat{R}}, \tag{10}$$

$$\hat{R} = \frac{\widehat{Err}^{(1)} - \overline{err}}{\hat{\gamma} - \overline{err}}, \tag{11}$$

$$\hat{\gamma} = \sum_{i=1}^{n}\sum_{j=1}^{n} Q[y_i, r_x(t_j)]/n^2. \tag{12}$$

Występująca w liczniku wzoru (11) miara przeuczenia $\widehat{Err}^{(1)} - \overline{err}$ skalowana jest z wykorzystaniem poziomu tzw. błędu braku informacji $\hat{\gamma}$ (ang. *no-information error rate*), który pojawiłby się wtedy, gdyby t_i oraz y_i były niezależne (wzór 12). W przypadku dychotomicznego problemu klasyfikacji błąd ten można przedstawić za pomocą zależności:

$$\hat{\gamma} = \hat{p}_1(1 - \hat{q}_1) + \hat{q}_1(1 - \hat{p}_1), \tag{13}$$

gdzie:
$\hat{p}_1$ – oznacza zaobserwowaną częstość y_i przyjmujących wartość 1,
$\hat{q}_1$ – oznacza zaobserwowaną częstość $r_x(t_j)$ przyjmujących wartość 1.

Uzyskany w ten sposób względny poziom przeuczenia $\hat{R}$, przyjmuje wartości z przedziału od 0 (dla braku przeuczenia, gdy $\widehat{Err}^{(1)} = \overline{err}$), do wartości 1 (kiedy poziom przeuczenia odpowiada wartościowo poziomowi $\hat{\gamma} - \overline{err}$). Waga $\hat{w}$ we wzorze (9) może przyjmować wówczas wartości z przedziału 0,632 do 1, a tym samym wartość estymatora $\widehat{Err}^{(0,632+)}$ jest nie mniejsza od $\widehat{Err}^{(0,632)}$ oraz nie większa od $\widehat{Err}^{(1)}$.

Efron, Tibshirani (1997) podkreślają w swojej pracy, iż mogą się zdarzyć sytuacje, gdy $\hat{\gamma} \leq \overline{err}$ lub $\overline{err} < \hat{\gamma} \leq \widehat{Err}^{(1)}$. Zaproponowali, by wówczas zmodyfikować wcześniej przedstawione zależności (wzory 7 i 11) w następujący sposób:

$$\widehat{Err}^{(1)\prime} = \min(\widehat{Err}^{(1)}, \hat{\gamma}), \tag{14}$$

$$\hat{R}' = \begin{cases} \dfrac{\widehat{Err}^{(1)} - \overline{err}}{\hat{\gamma} - \overline{err}} & jeżeli\ \widehat{Err}^{(1)} > \overline{err}\ i\ \hat{\gamma} > \overline{err}, \\ 0 & w\ pozostałych\ przypadkach. \end{cases} \tag{15}$$

W badaniach empirycznych wykorzystano oba opisane estymatory prawdziwego błędu predykcji (wzory 8 i 9). Należy jednak zaznaczyć, iż istnieje szereg innych, opisanych w literaturze metod repróbkowania służących do szacowania prawdziwego błędu predykcji (np. Jiang, Simon, 2007). Estymatory $\widehat{Err}^{(0,632)}$ oraz $\widehat{Err}^{(0,632+)}$ zostały wybrane przez autora ze względu na fakt, iż są one najbardziej popularne w literaturze.

3. WYKORZYSTANA PRÓBA BADAWCZA

Analiza porównawcza metod szacowania błędu predykcji klasyfikatora wymagała zgromadzenia odpowiedniej próby badawczej, reprezentującej dwie rozłączne grupy obiektów. W tym celu wykorzystano dane finansowe przedsiębiorstw o złej oraz dobrej kondycji finansowej. Kryterium decydującym o zaklasyfikowaniu przedsiębiorstw do pierwszej grupy był fakt ogłoszenia przez odpowiedni sąd ich upadłości. W celu wyselekcjonowania próby wykorzystano informacje zawarte w Internetowym Monitorze Sądowym i Gospodarczym. Postanowiono ograniczyć selekcję przedsiębiorstw do spółek akcyjnych, reprezentujących trzy różne branże gospodarki. W ten sposób, biorąc także pod uwagę dostępność danych finansowych, wyselekcjonowano:

– 30 spółek akcyjnych z branży budownictwo (PKD 41.10-43.99z),
– 30 spółek akcyjnych z branży przetwórstwo przemysłowe (PKD 10.11-33.20z),
– 30 spółek akcyjnych z branży handel hurtowy i detaliczny (PKD 46.11-47.99z).

Do każdej z nich została dobrana spółka akcyjna o dobrej kondycji finansowej. Za kryteria dopasowania poszczególnych par przyjęto: sektor, działalność główną oraz wielkość aktywów. Dane finansowe spółek upadłych pochodziły z ich sprawozdań finansowych z roku, poprzedzającego ten w którym złożono pierwszy wniosek o ogłoszenie upadłości. Dotyczyły one lat 2000–2013. Sprawozdania finansowe dla spółek zdrowych pochodziły z tych samym okresów. Źródłem danych były bazy firm Notoria Serwis i Bisnode Dun & Bradstreet oraz Monitor Polski B.

Tabela 1.

Lista wskaźników finansowych wykorzystanych w badaniu

Symbol	Formuła wskaźnika
Wskaźniki rentowności	
ROA	zysk netto / średnia wartość aktywów
ROE	zysk netto / średnia wartość kapitałów własnych
ZB	zysk brutto / średnia wartość aktywów
ZS	zysk ze sprzedaży / średnia wartość aktywów
MZ	zysk brutto / przychody ze sprzedaży
MZ2	zysk netto / przychody ze sprzedaży
MZO	zysk operacyjny / przychody ze sprzedaży

Tabela 1. (cd.)

Symbol	Formuła wskaźnika
Wskaźniki płynności	
KP	kapitał pracujący / suma bilansowa
WBP	majątek obrotowy / zobowiązania krótkoterminowe
WSP	(majątek obrotowy-zapasy)/zobowiązania krótkoterminowe
WPP	(majątek obrotowy-zapasy-należności)/zobowiązania krótkoterminowe
Wskaźniki struktury kapitałowo-majątkowej	
ZO	zobowiązania ogółem / aktywa ogółem
ZD	zobowiązania długoterminowe / aktywa ogółem
KW	kapitał własny / aktywa ogółem
KWZ	kapitał własny / zobowiązania ogółem
Wskaźniki sprawności działania	
RN	średnia wartość należności / przychody ze sprzedaży netto*365
RZ	średnia wartość zapasów / przychody ze sprzedaży netto*365
RZob	średnia wartość zobowiązań / przychody ze sprzedaży*365
Rakt	średnia wartość aktywów/przychody ze sprzedaży*365

Źródło: opracowanie własne.

W badaniach empirycznych wykorzystano 19 wskaźników finansowych charakteryzujących rentowność, płynność, strukturę kapitałowo-majątkową oraz sprawność działania przedsiębiorstw (tabela 1).

Ich wyboru dokonano na podstawie przeglądu literatury – są to wskaźniki pojawiające się najczęściej w modelach prognozowania upadłości. W wyborze kierowano się także dostępnością danych w sprawozdaniach finansowych spółek. Wartości wskaźników finansowych zostały obliczone dla roku poprzedzającego ten, w którym złożono pierwszy wniosek o ogłoszenie upadłości przedsiębiorstwa – dla spółek, wobec których ogłoszono upadłość oraz dla tego samego roku dla odpowiadających im spółek zdrowych.

Wśród założeń przyjmowanych przy konstruowaniu funkcji dyskryminacyjnej są te dotyczące rozkładu normalnego oraz równości wariancji zmiennych opisujących obiekty w badanych grupach. W literaturze podkreśla się jednak, iż niespełnienie tych wymagań nie pogarsza istotnie wyników uzyskanych za pomocą omawianej metody (Hand, 1981, s. 27; Hadasik, 1998, s. 94). Z tego też powodu – pomimo niespełnienia wspomnianych założeń[2] – w badaniu wykorzystano liniową analizę dyskryminacyjną.

[2] Jedynie w przypadku trzech wskaźników (*KP*, *ZO* oraz *KW*) dla spółek zdrowych oraz jednego wskaźnika (*WBP*) dla spółek, wobec których ogłoszono upadłość, spełnione jest założenie o ich rozkładzie normalnym. Tylko w przypadku trzech wskaźników sprawności działania (*RN*, *RZ* oraz *Rakt*) nie

4. EMPIRYCZNA ANALIZA PORÓWNAWCZA WYBRANYCH METOD SZACOWANIA BŁĘDU PREDYKCJI KLASYFIKATORA

Celem badania empirycznego jest porównanie różnych metod szacowania prawdziwego błędu predykcji. Przeprowadzona analiza składa się z M iteracji. W każdym powtórzeniu m ($m = 1, 2, \ldots, M$) następuje, spośród przedstawionej grupy 180 spółek akcyjnych, losowanie stratyfikowanej[3] podpróby, składającej się z n obiektów. Każda z nich pełni w badaniu dwojaką rolę. Po pierwsze, na jej podstawie zostają oszacowane, zgodnie z przedstawionymi wcześniej metodami, estymatory prawdziwego błędu predykcji $\widehat{Err}$. Po drugie, traktując każdą z wylosowanych podprób jako próbę uczącą, a pozostałe $180 - n$ obiektów jako dużą, niezależną próbę testową, oszacowano wartość prawdziwego błędu predykcji Err (zgodnie z wzorem (2)). W celu porównania różnych metod szacowania wykorzystano trzy następujące miary: odchylenie standardowe (*SD*) zmiennej $\widehat{Err}$ oraz obciążenie (*Bias*) i błąd średniokwadratowy (*MSE*) zmiennej $\widehat{Err} - Err$ (wzory 16–18).

$$SD = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(\widehat{Err}_{n,m} - \overline{\widehat{Err}_n}\right)^2}, \tag{16}$$

$$Bias = \frac{1}{M}\sum_{m=1}^{M}\left(\widehat{Err}_{n,m} - Err_{n,m}\right), \tag{17}$$

$$MSE = \frac{1}{M}\sum_{m=1}^{M}\left(\widehat{Err}_{n,m} - Err_{n,m}\right)^2. \tag{18}$$

W przypadku każdej metody szacowania prawdziwego błędu predykcji, proces uczenia poprzedzony był każdorazowo selekcją zmiennych do modelu. Inne podejście, polegające na tylko jednorazowym wyborze zmiennych do modelu, skutkowałoby obciążonym estymatorem prawdziwego błędu predykcji (Simon i inni, 2003; Jiang, Simon, 2007).

W celu przeprowadzenia badania empirycznego przyjęto następujące założenia:

– liczba iteracji M równa jest 1000,
– liczba obiektów losowanych do podprób w kolejnych iteracjach $n = 60$, $n = 90$, $n = 120$,

---

ma podstaw do odrzucenia hipotezy zerowej mówiącej o tym, iż populacje, z których pochodzą obiekty mają jednakową wariancję.

[3] Każda wylosowana podpróba charakteryzuje się taką samą proporcją spółek, które ogłosiły upadłość i spółek zdrowych, jak miało to miejsce w oryginalnej próbie 180 spółek, czyli 1:1.

– dla prostej metody podziału przyjęto proporcje p podziału próby na uczącą oraz testową $p = \frac{1}{5}$, $p = \frac{1}{3}$, $p = \frac{1}{2}$.
– dla k-krotnej walidacji krzyżowej przyjęto $k = n$, $k = 3$, $k = 5$.
– dla powtarzanej k-krotnej walidacji krzyżowej oraz powtarzanej prostej metody podziału przyjęto liczbę powtórzeń równą 10,
– dla metod wielokrotnego repróbkowania przyjęto liczebność losowanych prób typu bootstrap $B = 50$.

Listę badanych estymatorów błędu predykcji przedstawiono w tabeli 2.

Tabela 2.

Estymatory błędu predykcji porównywane w badaniu

Oznaczenie estymatora	Metoda szacowania
RS	metoda resubstytucji – próba testowa odpowiada próbie uczącej
Split 1/5	metoda prosta podziału (próba testowa – 1/5 całej próby)
Split 1/3	metoda prosta podziału (próba testowa – 1/3 całej próby)
Split 1/2	metoda prosta podziału (próba testowa – 1/2 całej próby)
CV3	walidacja krzyżowa – podział próby na 3 części
CV5	walidacja krzyżowa – podział próby na 5 części
CV3 r10	powtarzana walidacja krzyżowa – podział próby na 3 części, 10 powtórzeń
CV5 r10	powtarzana walidacja krzyżowa – podział próby na 5 części, 10 powtórzeń
LOOCV	walidacja krzyżowa typu „pozostaw jedną poza”
0,632	wielokrotne repróbkowanie, estymator 0,632, liczba losowanych prób typu bootstrap $B = 50$
0,632+	wielokrotne repróbkowanie, estymator 0,632+, liczba losowanych prób typu bootstrap $B = 50$
rSplit	powtarzana 50-krotna prosta metoda podziału (próba testowa – 1/5 całej próby)

Źródło: opracowanie własne.

Do konstrukcji modelu klasyfikacyjnego (klasyfikatora) wykorzystano, jak wspomniano wcześniej, liniową analizę dyskryminacyjną. W celu uniknięcia skorelowania zmiennych opisujących obiekty przyjęto, iż każdorazowo przed procesem uczenia usuwane są te zmienne, które są silnie skorelowane z pozostałymi (współczynnik korelacji Pearsona wyższy od 0,90)[4]. Dodatkowo w celu wyselekcjonowania tych wskaźników finansowych, które charakteryzują się najwyższą zdolnością dyskryminacyjną w badaniu wykorzystano 3 statystyczne metody wyboru zmiennych:

[4] W przypadku gdy dwie zmienne są silnie skorelowane z dalszej analizy usuwana jest ta z nich, dla której średnia z wartości bezwzględnych współczynników korelacji między tą zmienną z pozostałymi jest wyższa.

– wybór 4 zmiennych, których wartość bezwzględna statystyki *t*-studenta dla testu porównującego średnią wartość zmiennej dla badanych grup jest najwyższa (metoda oznaczona dalej jako *t4zmienne*),
– metoda krokowa w przód, przyjęto poziom istotności dla wartości statystyki F równy 0,1 (*krokowa*),
– dobór zmiennych silnie skorelowanych ze zmienną grupującą – współczynnik korelacji Pearsona jest statystycznie istotny przy poziomie $\alpha = 0{,}05$ (*korelacja*).

Całość obliczeń została wykonana z wykorzystaniem środowiska statystycznego R.

W tabelach 3–5 przedstawiono charakterystyki tj. odchylenie standardowe (*SD*), obciążenie (*Bias*) oraz błąd średniokwadratowy (*MSE*) estymatorów błędu predykcji, uzyskane dla wymienionych statystycznych metod wyboru zmiennych do modelu.

Analizując dane zawarte w tabelach można zauważyć, iż w przypadku estymatora resubstytucji analizowana zmienna $\widehat{Err} - Err$, bez względu na liczebność losowanych w kolejnych iteracjach podprób, zawsze charakteryzuje się ujemnym obciążeniem. W przypadku estymatorów uzyskanych prostą metodą podziału odchylenie standardowe zmiennej $\widehat{Err}$ jest zdecydowanie wyższe, od pozostałych porównywanych w badaniu. Zestawiając te wartości z błędami predykcji oszacowanymi za pomocą metod walidacji krzyżowej wyraźnie widać, iż te drugie charakteryzują się zarówno znacznie mniejszym odchyleniem standardowym, jak i mniejszymi wartościami błędów średniokwadratowych. Biorąc pod uwagę obciążenie analizowanej zmiennej, najlepsze własności (wśród estymatorów szacowanych metodą jednokrotnej walidacji krzyżowej) posiada estymator CV5.

Tabela 3.

Analiza porównawcza estymatorów błędu predykcji. Metoda wyboru zmiennych do modelu: *t4zmienne*

Estymator	$n = 60$			$n = 90$			$n = 120$		
	SD	Bias	MSE	SD	Bias	MSE	SD	Bias	MSE
RS	0,0600	-0,0317	0,0046	0,0549	-0,0232	0,0036	0,0581	-0,0189	0,0037
Split 1/5	0,1283	-0,0041	0,0165	0,1098	-0,0047	0,0121	0,0981	-0,0010	0,0096
Split 1/3	0,1036	-0,0073	0,0108	0,0868	-0,0048	0,0076	0,0852	-0,0066	0,0073
Split 1/2	0,0918	-0,0189	0,0088	0,0825	-0,0140	0,0070	0,0803	-0,0119	0,0066
CV3	0,0702	-0,0082	0,0050	0,0632	-0,0066	0,0040	0,0639	-0,0074	0,0041
CV5	0,0671	-0,0036	0,0045	0,0607	-0,0042	0,0037	0,0630	-0,0037	0,0040
CV3r10	0,0625	-0,0087	0,0040	0,0570	-0,0072	0,0033	0,0610	-0,0063	0,0038
CV5r10	0,0621	-0,0046	0,0039	0,0569	-0,0041	0,0033	0,0606	-0,0028	0,0037
LOOCV	0,0724	-0,0163	0,0055	0,0601	-0,0111	0,0037	0,0625	-0,0072	0,0040
0,632	0,0592	-0,0021	0,0035	0,0548	-0,0004	0,0030	0,0594	-0,0019	0,0035
0,632+	0,0598	-0,0003	0,0035	0,0551	-0,0011	0,0030	0,0591	-0,0008	0,0035
rSplit	0,0645	-0,0053	0,0042	0,0575	-0,0044	0,0033	0,0617	-0,0027	0,0038

Źródło: opracowanie własne.

Tabela 4.

Analiza porównawcza estymatorów błędu predykcji. Metoda wyboru zmiennych do modelu: *krokowa*

Estymator	$n = 60$			$n = 90$			$n = 120$		
	SD	Bias	MSE	SD	Bias	MSE	SD	Bias	MSE
RS	0,0615	-0,0494	0,0062	0,0573	-0,0399	0,0049	0,0605	-0,0376	0,0051
Split 1/5	0,1314	-0,0083	0,0173	0,1097	-0,0061	0,0121	0,1008	-0,0001	0,0102
Split 1/3	0,1074	-0,0181	0,0119	0,0888	-0,0077	0,0079	0,0831	-0,0042	0,0069
Split 1/2	0,0953	-0,0246	0,0097	0,0813	-0,0182	0,0069	0,0794	-0,0090	0,0064
CV3	0,0716	-0,0129	0,0053	0,0647	-0,0081	0,0043	0,0652	-0,0020	0,0043
CV5	0,0683	-0,0082	0,0047	0,0633	-0,0052	0,0040	0,0649	-0,0006	0,0042
CV3r10	0,0590	-0,0132	0,0036	0,0557	-0,0083	0,0032	0,0610	-0,0033	0,0037
CV5r10	0,0607	-0,0080	0,0038	0,0562	-0,0047	0,0032	0,0611	-0,0012	0,0037
LOOCV	0,0743	-0,0221	0,0060	0,0680	-0,0160	0,0049	0,0679	-0,0097	0,0047
0,632	0,0581	-0,0006	0,0034	0,0551	-0,0019	0,0030	0,0600	-0,0060	0,0036
0,632+	0,0587	-0,0047	0,0035	0,0554	-0,0013	0,0031	0,0603	-0,0008	0,0036
rSplit	0,0614	-0,0083	0,0038	0,0570	-0,0052	0,0033	0,0621	-0,0017	0,0039

Źródło: opracowanie własne.

Tabela 5.

Analiza porównawcza estymatorów błędu predykcji. Metoda wyboru zmiennych do modelu: *korelacja*

Estymator	$n = 60$			$n = 90$			$n = 120$		
	SD	Bias	MSE	SD	Bias	MSE	SD	Bias	MSE
RS	0,0631	-0,0730	0,0093	0,0558	-0,0532	0,0060	0,0609	-0,0431	0,0056
Split 1/5	0,1333	-0,0110	0,0179	0,1063	-0,0065	0,0114	0,1017	-0,0057	0,0104
Split 1/3	0,1065	-0,0099	0,0114	0,0884	-0,0108	0,0079	0,0850	-0,0084	0,0073
Split 1/2	0,1011	-0,0229	0,0108	0,0815	-0,0232	0,0072	0,0801	-0,0174	0,0067
CV3	0,0705	-0,0136	0,0052	0,0649	-0,0119	0,0044	0,0657	-0,0100	0,0044
CV5	0,0730	-0,0077	0,0054	0,0601	-0,0056	0,0036	0,0642	-0,0057	0,0042
CV3r10	0,0620	-0,0144	0,0041	0,0564	-0,0118	0,0033	0,0618	-0,0095	0,0039
CV5r10	0,0630	-0,0073	0,0040	0,0563	-0,0063	0,0032	0,0618	-0,0052	0,0038
LOOCV	0,0727	-0,0185	0,0056	0,0620	-0,0126	0,0040	0,0649	-0,0117	0,0044
0,632	0,0599	-0,0070	0,0036	0,0538	-0,0045	0,0029	0,0603	-0,0046	0,0037
0,632+	0,0611	-0,0021	0,0037	0,0543	-0,0005	0,0029	0,0606	-0,0015	0,0037
rSplit	0,0647	-0,0085	0,0043	0,0566	-0,0069	0,0033	0,0620	-0,0051	0,0039

Źródło: opracowanie własne.

Prezentując teoretyczne podstawy różnych metod szacowania błędu predykcji zwrócono uwagę na to, iż błędy predykcji estymowane metodami (jednokrotnej) walidacji krzyżowej „narażone są" na dodatkową zmienność, określaną mianem wewnętrznej. Przedstawionym sposobem jej redukcji było wielokrotne powtórzenie procesu szacowania błędu predykcji – powtarzana *k*-krotna walidacja krzyżowa. Weryfikując skuteczność takiego postępowania, należy porównać charakterystyki estymatorów CV3 oraz CV5 z ich odpowiednikami, uwzględniającymi 10-krotne powtarzanie procesu szacowania błędu: CV3r10 oraz CV5r10. Analizując wyniki zawarte w tabelach można stwierdzić, iż uwzględnienie powtórzeń w procesie szacowania błędu predykcji w każdym przypadku redukuje, zgodnie z oczekiwaniami, zmienność wyników uzyskanych z wykorzystaniem badanych estymatorów. Ma to swoje odzwierciedlenie w zmniejszonych wartościach błędu średniokwadratowego. Na tej podstawie można wyciągnąć wniosek, iż większy koszt obliczeniowy, związany z 10-krotnym powtarzaniem procesu szacowania błędu predykcji, pozwala osiągnąć korzystniejsze własności uzyskanych estymatorów.

Analiza danych zawartych w tabelach 3–5 pozwala stwierdzić, iż niezależnie od przyjętej metody wyboru zmiennych do modelu oraz wielkości wylosowanej podpróby, wartości odchylenia standardowego oraz błędu średniokwadratowego są najniższe w przypadku estymatorów uzyskanych metodami wielokrotnego repróbkowania (oznaczone jako 0,632 oraz 0,632+). Obciążenie analizowanej zmiennej, w przypadku estymatora 0,632 (wzór 8) jest zawsze najniższe, jednak bardzo często ujemne. Potwierdzono tym samym, o czym wspomniano w pierwszym podrozdziale, iż estymator wielokrotnego repróbkowania 0,632 (wzór 8) niedoszacowuje wartości błędu predykcji.

5. PODSUMOWANIE I WNIOSKI

Celem niniejszego artykułu były przegląd oraz empiryczna analiza porównawcza wybranych metod szacowania błędu predykcji klasyfikatora, skonstruowanego z wykorzystaniem liniowej analizy dyskryminacyjnej. Badanie empiryczne zostało przeprowadzone na przykładzie problemu prognozowania upadłości spółek akcyjnych w Polsce.

W przeprowadzonej analizie zbadano własności statystyczne 12 estymatorów prawdziwego błędu predykcji klasyfikatora (ang. *true error rate*). Podsumowując wyniki przeprowadzonego badania sformułować można następujące wnioski, dotyczące metod szacowania błędu predykcji klasyfikatora:

– Estymator resubstytucji ma znaczącą tendencję do niedoszacowywania wartości prawdziwego błędu predykcji. Uzyskane za jego pomocą wyniki są nadmiernie optymistyczne.
– Estymatory uzyskane za pomocą prostych metod podziału charakteryzują się najwyższą zmiennością spośród wszystkich poddanych analizie.
– Uwzględnienie powtórzeń w metodach walidacji krzyżowej poprawia jakość uzyskanych za ich pomocą estymatorów. Związane jest to jednak z wyższym kosztem obliczeniowym.

– Estymatory 0,632 szacowane metodą wielokrotnego repróbkowania charakteryzują się w większości przypadków ujemnym obciążeniem, co oznacza, że niedoszacowują one wartości prawdziwego błędu predykcji.
– Metoda wyboru zmiennych do modelu oraz liczebność losowanych podprób n nie wpływa na wnioski, jakie można sformułować na podstawie analizy porównawczej badanych estymatorów.

Podsumowując, uzyskane wyniki świadczą o tym, iż estymator prawdziwego błędu predykcji 0,632+ charakteryzuje się najbardziej pożądanymi własnościami w przypadku prognozowania upadłości spółek akcyjnych w Polsce.

Szacowanie błędu predykcji jest kluczowym etapem konstrukcji każdego klasyfikatora. Z tego też powodu, należy kontynuować przedstawiony kierunek badań. W badaniu wykorzystano wybrane estymatory punktowe. W ostatnich latach w literaturze podkreśla się, iż uzyskane w ten sposób wyniki, szczególnie w przypadku małolicznych prób badawczych, mogą być niewiarygodne. Z tego też powodu wskazuje się, iż lepszym, choć też nie idealnym, rozwiązaniem może być określanie przedziałów ufności dla szacowanych błędów predykcji klasyfikatora (Hanczar, Dougherty, 2013).

LITERATURA

Braga-Neto U. M., Dougherty E. R., (2004), Is Cross-validation for Small-sample Microarray Classification?, *Bioinformatics*, 20 (3), 374–380.

Efron B., (1983), Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation, *Journal of the American Statistical Association*, 78 (382), 316–331.

Efron B., Tibshirani R. J., (1997), Improvements on Cross-Validation: The .632+ Bootstrap Method, *Journal of the American Statistical Association*, 92 (438), 548–560.

Gatnar E., (2001), *Nieparametryczna metoda dyskryminacji i regresji*, Wydawnictwo Naukowe PWN, Warszawa.

Gatnar E., (2008), *Podejście wielomodelowe w zagadnieniach dyskryminacji i regresji*, Wydawnictwo Naukowe PWN, Warszawa.

Geisser S., (1975), The Predictive Sample Reuse Method With Applications, *Journal of the American Statistical Association*, 70, 320–328.

Hadasik D., (1998), *Upadłość przedsiębiorstw w Polsce i metody jej prognozowania*, Zeszyty naukowe – seria II, Prace habilitacyjne, Zeszyt 153, Akademia Ekonomiczna w Poznaniu, Poznań.

Hanczar B., Dougherty E. R., (2013), The Reliability of Estimated Confidence Intervals for Classification Error Rates When Only a Single Sample is Available, *Pattern Recognition*, 46, 1067–1077.

Hand D. J., (1981), *Discrimination and Classification*, John Wiley & Sons, Chichester.

Isaksson A., Wallman M., Goransson H., Gustafsson M. G., (2008), Cross-Validation and Bootstrapping are Unreliable in Small Sample Classification, *Pattern Recognition*, 29, 1960–1965.

Jiang W., Simon R., (2007), A Comparison of Bootstrap Methods and an Adjusted Bootstrap Approach for Estimating Prediction Error in Microarray Classification, *Statistics in Medicine*, 26, 5320–5334.

Kim J. H., (2009), Estimating Classification Error Rate: Repeated Cross-Validation, Repeated Hold-Out and Bootstrap, *Computational Statistics and Data Analysis*, 53, 3735–3745.

Lachenbruch P. A., Mickey M. R., (1968), Estimation of Error Rates in Discriminant Analysis, *Technometrics*, 10, 1–11.

McLachlan G. J., (1992), *Discriminant Analysis and Statistical Pattern Recognition*, John Wiley & Sons, Inc.

Molinaro A. M., Simon R., Pfeiffer R. M., (2005), Prediction Error Estimation: A Comparison of Resampling Methods, *Bioinformatics*, 21, 3301–3307.

Ripley B. D., (1996), *Pattern Recognition and Neural Networks*, Cambrige University Press.

Simon R., Radmacher M. D., Dobbin K., McShane L. M., (2003), Pitfalls in the Use of DNA Microarray Data for Diagnostic and Prognostic Classification, *Journal of the National Cancer Institute*, 95 (1), 14–18.

Wehberg S., Schumacher M., (2004), A Comparison of Nonparametric Error Rate Estimation Methods in Classification Problems, *Biometrical Journal*, 46, 35–47.

ANALIZA PORÓWNAWCZA WYBRANYCH METOD SZACOWANIA BŁĘDU PREDYKCJI KLASYFIKATORA

Streszczenie

Klasyfikacją nazywamy algorytm postępowania, który przydziela badane obserwacje/obiekty, bazując na ich cechach do określonych populacji. W tym celu konstruowany jest odpowiedni model – klasyfikator. Miarą jego jakości jest przede wszystkim zdolność predykcyjna, mierzona m.in. za pomocą prawdziwego błędu predykcji. Wartość tego błędu, ze względu na brak odpowiednio dużej, niezależnej próby testowej, musi być często szacowana na podstawie dostępnej próby uczącej.

Celem artykułu jest dokonanie przeglądu oraz empirycznej analizy porównawczej wybranych metod szacowania błędu predykcji klasyfikatora, skonstruowanego z wykorzystaniem liniowej analizy dyskryminacyjnej. Zbadano, czy wyniki analizy uzależnione są od wielkości próby oraz metody wyboru zmiennych do modelu. Badanie empiryczne zostało przeprowadzone na przykładzie problemu prognozowania upadłości spółek akcyjnych w Polsce.

Słowa kluczowe: błąd predykcji, walidacja krzyżowa, prosta metoda podziału, wielokrotne reprób-kowanie, upadłość przedsiębiorstw, klasyfikacja

COMPARATIVE ANALYSIS OF SELECTED METHODS FOR ESTIMATING THE PREDICTION ERROR OF CLASSIFIER

Abstract

Classification is an algorithm, which assigns studied companies, taking into consideration their attributes, to specific population. An essential part of it is classifier. Its measure of quality is especially predictability, measured by true error rate. The value of this error, due to lack of sufficiently large and independent test set, must be estimated on the basis of available learning set.

The aim of this article is to make a review and compare selected methods for estimating the prediction error of classifier, constructed with linear discriminant analysis. It was examined if the results of the analysis depends on the sample size and the method of selecting variables for a model. Empirical research was made on example of problem of bankruptcy prediction of join-stock companies in Poland.

Keywords: prediction error, cross-validation, holdout method, bootstrapping, corporate bankruptcy, classification

Redakcja Przeglądu Statystycznego składa podziękowania Recenzentom artykułów nadesłanych do publikacji w roku 2016 w osobach:

Jacek Batóg
Andrzej Bąk
Jakub Bijak
Joanna Bruzda
Ryszard Doman
Kamil Fijorek
Michał Gradzewicz
Jakub Growiec
Montserrat Guillén
Tom Hyclak
Antonio Irpino
Bogumił Kamiński
Robert Kelm
Paweł Kliber
Michał Konopczyński
Jadwiga Kostrzewska
Maciej Kostrzewski
Stanisław Maciej Kot
Karol Kukuła
Ireneusz Kuropka
Tomasz Kuszewski
Jacek Kwiatkowski
Łukasz Kwiatkowski
Joanna Landmesser
Jacek Leśkow
Agnieszka Lipieta
Michał Majsterek
Maciej Malaczewski
Krzysztof Malaga
Marta Małecka
Małgorzata Markowska
Jerzy Marzec
Alina Matei
Elżbieta Mączyńska-Ziemacka
Władysław Milo
Paweł Miłobędzki
Krzysztof Najman
Monique Noirhomme
Witold Orzeszko
Jorge Oviedo
Anna Pajor
Emil Panek
Tomasz Panek
Barbara Pawełek
Krzysztof Piasecki
Krzysztof Piontek
Mateusz Pipień
Piotr Płuciennik
Michał Ramsza
Sylwia Roszkowska
Małgorzata Rószkiewicz
Dobromił Serwa
Ewa Soja
Honorata Sosnowska
Józef Stawicki
Mirosław Szreder
Tomasz Tokarski
Grażyna Trzpiot
Wit Urban
Marek Walesiak
Markus Walzl
Ewa Wędrowska
Dariusz Wędzki
Feliks Wysocki
Anna Zamojska
Henryk Zawadzki