

METODY ILOŚCIOWE
W BADANIACH EKONOMICZNYCH

QUANTITATIVE METHODS
IN ECONOMICS

Vol. XIV, No. 3

Warsaw University of Life Sciences – SGGW
Faculty of Applied Informatics and Mathematics
Department of Econometrics and Statistics

METODY ILOŚCIOWE
W BADANIACH EKONOMICZNYCH

**QUANTITATIVE METHODS
IN ECONOMICS**

Volume XIV, No. 3

Warsaw 2013

EDITORIAL BOARD

Zbigniew Binderman, Bolesław Borkowski – Editor-in-Chief
Hanna Dudek – Managing Editor, Joanna Landmesser, Wojciech Zieliński

SCIENTIFIC BOARD

Zbigniew Binderman (Warsaw University of Life Sciences – SGGW, Poland)
Paolo Gajo (University of Florence, Italy)
Evgeny Grebenikov (Computing Centre of Russia Academy of Sciences, Moscow, Russia)
Yuriy Kondratenko (Black Sea State University, Ukraine)
Vassilis Kostoglou (Alexander Technological Educational Institute of Thessaloniki, Greece)
Robert Kragler (University of Applied Sciences, Weingarten, Germany)
Yochanan Shachmurove (The City College of The City University of New York, USA)
Alexander N. Prokopenya (Brest University, Belarus)
Ewa Marta Syczewska (Warsaw School of Economics, Poland)
Andrzej Wiatrak (University of Warsaw, Poland)
Dorota Witkowska (Warsaw University of Life Sciences – SGGW, Poland)
Monika Krawiec – Secretary (Warsaw University of Life Sciences – SGGW, Poland)

TECHNICAL EDITORS

Jolanta Kotlarska, Elżbieta Saganowska

LIST OF REVIEWERS

Wiktor Adamus, Iwona Bąk, Aneta Becker, Jarosław Becker, Jacek Bednarz,
Lucyna Błażejczyk–Majka, Ryszard Budziński, Ludosław Drelichowski, Szczepan Figiel,
Paolo Gajo, Stanisław Gędek, Henryk Gurgul, Stanisław Kasiewicz, Joanna Kisielińska,
Yuriy Kondratenko, Stanisław Kot, Vassilis Kostoglou, Barbara Kowalczyk, Leszek Kuchar,
Tadeusz Kufel, Karol Kukuła, Ryszard Kutner, Tadeusz Kwater, Waclaw Laskowski,
Wanda Marcinkowska–Lewandowska, Kesra Nermend, Magdalena Osińska, Maria Parlińska,
Marian Podstawka, Artur Prędki, Alexander N. Prokopenya, Włodzimierz Rembisz,
Yochanan Shachmurove, Ewa Marta Syczewska, Stanisław Stańko, Jacek Strojny,
Michał Świtłyk, Beata Pułaska–Turyna, Tadeusz Waściński, Andrzej Wiatrak,
Antoni Wiliński, Bartosz Witkowski, Aldon Zalewski, Michał Zasada, Wojciech Zięta

Language Editor: Agata Kropiwek

Statistical Editor: Wojciech Zieliński

Native speaker: Yochanan Shachmurove

Subject Editors:

Econometrics & Statistics – Bolesław Borkowski

Financial Engineering – Dorota Witkowska

Multidimensional Data Analysis – Wiesław Szczesny

Mathematical Economy – Zbigniew Binderman

ISSN 2082 – 792X

© Copyright by Katedra Ekonometrii i Statystyki SGGW

Warsaw 2013, Volume XIV, No. 3

The original version is the paper version

Journal homepage: qme.sggw.pl

Published by Warsaw University of Life Sciences Press

SPIS TREŚCI

Iwona Bąk – Badanie przestrzennego zróżnicowania poziomu ekoturystyki w Polsce z wykorzystaniem analizy dyskryminacyjnej	7
Agata Binderman – Dynamika regionalnego zróżnicowania rolnictwa w Polsce w latach 1998-2010	17
Zbigniew Binderman, Bolesław Borkowski, Wiesław Szczesny – Zastosowanie metryki Minkowskiego do pomiaru zmian koncentracji	27
Mariusz Franciszek Kaszubowski – Modele płac typu Mincera na przykładzie pracowników naukowo-dydaktycznych polskich wyższych uczelni publicznych	39
Joanna Małgorzata Landmesser – Dekompozycja różnic pomiędzy kobietami i mężczyznami w procesie opuszczania stanu bezrobocia	51
Piotr Łukasiewicz, Krzysztof Karpio, Arkadiusz Orłowski – Dochody indywidualne a dochody gospodarstw domowych w Polsce w latach 2000 – 2010	62
Jacek Strojny – Eksport a dynamika produkcji sektorów rolnych krajów UE	71
Maria Szmuksta-Zawadzka, Jan Zawadzki – Modele harmoniczne ze złożoną sezonowością w prognozowaniu szeregów czasowych z lukami systematycznymi	81
Jadwiga Zaród – Wielokryterialny model optymalizacyjny przeciętnego gospodarstwa rolnego w województwie zachodniopomorskim	91
Krzysztof Zmarzłowski, Grzegorz Koszela – Czynniki różnicujące spożycie wyrobów spirytusowych w polskich gospodarstwach domowych w latach 1999 – 2008	101

BADANIE PRZESTRZENNEGO ZRÓŻNICOWANIA POZIOMU EKOTURYSTYKI W POLSCE Z WYKORZYSTANIEM ANALIZY DYSKRYMINACYJNEJ

Iwona Bąk

Katedra Zastosowań Matematyki w Ekonomii
Zachodniopomorski Uniwersytet Technologiczny w Szczecinie
e-mail: iwona.bak@zut.edu.pl

Streszczenie: Celem artykułu jest analiza przestrzennego zróżnicowania poziomu atrakcyjności podregionów w Polsce z punktu widzenia możliwości rozwoju w nich turystyki przyjaznej środowisku przyrodniczemu, tzw. ekoturystyki. Do analizy wykorzystano wskaźniki charakteryzujące atrakcyjność środowiska naturalnego podregionów (stymulanty) oraz wskaźniki mierzące poziom jego zanieczyszczenia (destymulanty). Klasyfikacji podregionów dokonano za pomocą analizy dyskryminacyjnej. Wstępnej klasyfikacji obiektów na grupy, a tym samym wyboru zmiennej grupującej, dokonano stosując metodę k -średnich.

Słowa kluczowe: ekoturystyka, analiza dyskryminacyjna, metoda k -średnich

WSTĘP

Turystyka jako aktywność gospodarcza w istotny sposób wpływa na środowisko naturalne i antropogeniczne. Oprócz wielu pozytywnych aspektów coraz częściej zaznaczane są negatywne skutki jej rozwoju, głównie o charakterze środowiskowym. Masowa, nierównoważona i niekontrolowana turystyka przyczynia się do powstania wielu trwałych szkód we wszystkich komponentach przyrody. Wymusza to poszukiwanie oraz lansowanie takich rodzajów turystyki, które spowodują najmniejsze straty na obszarze recepcyjnym [Karczmarska 2010, s. 211]. Światowe i europejskie trendy zmierzają w kierunku turystyki zrównoważonej, która „dotyczy szeroko pojętej koncepcji rozwoju turystycznego przyjaznego środowisku w regionach wiejskich oraz miastach, w małych ośrodkach turystycznych oraz wielkich centrach rozrywkowych i wypoczynkowych itp. ...” [Zaręba 2000, s. 35]. Znacznie węższym pojęciem jest

ekoturystyka, będąca jedną z form podróżowania ściśle związaną z przyrodą i rdzenną kulturą obszarów o wysokich walorach naturalnych. Ekoturystyka (określana również w literaturze jako turystyka zielona – *green tourism*, przyrodnicza – *nature tourism* i ekoetnoturystyka – *ecoethnotourism*) posiada trzy bardzo ważne cechy wyróżniające ją spośród innych form podróżowania [Zaręba 2000, s. 48; Łobożewicz, Bieńczyk 2001, s. 125]:

1. jest formą aktywnego i dogłębnego zwiedzania obszarów o wybitnych walorach przyrodniczych i kulturowych,
2. strzeże harmonii ekosystemów przyrodniczych i odrębności kulturowej lokalnych społeczności,
3. dostarcza środków finansowych skutecznej ochronie wartości dziedzictwa przyrodniczego i kulturowego, przynosząc realne korzyści ekonomiczno-społeczne ludności miejscowej.

Zapotrzebowanie na tę formę turystyki rośnie, podobnie jak na inne rodzaje usług uważane za przyjazne środowisku. Zainteresowanie ekoturystyką szacuje się na poziomie 10-15% popytu turystycznego w skali ogólnoswiatowej. Skala ta różni się w zależności od kraju oraz poziomu edukacji ekologicznej społeczeństwa [Mika 2008, s. 209].

Również w Polsce obserwuje się nowe tendencje dotyczące rozwoju turystyki przyjaznej środowisku. Władze wielu gmin i miejscowości turystycznych wychodzą z założenia, że jedynie czyste środowisko i nieskażona przyroda mogą przyciągnąć turystów i zachęcić ich do ponownego odwiedzenia regionu. Inwestują więc w poprawę czystości wód i powietrza, tworzą szlaki turystyczne, dbają o szatę roślinną itp.

Przedmiotem artykułu jest badanie przestrzennego zróżnicowania poziomu atrakcyjności podregionów w Polsce z punktu widzenia możliwości rozwoju w nich turystyki przyjaznej środowisku przyrodniczemu. Atuty poszczególnych podregionów to: zróżnicowanie krajobrazu, ściśle powiązane z różnorodnością warunków klimatycznych i glebowych, urozmaicone formy ukształtowania powierzchni, różnaita szata roślinna. Z różnorodnością warunków terenu wiąże się możliwość uprawiania wielu form ekoturystyki.

METODA BADANIA

Do badania przestrzennego zróżnicowania rozwoju ekoturystyki w Polsce zastosowano metodę dyskryminacyjną. Istota jej polega na oszacowaniu jednowymiarowych liniowych funkcji obserwacji (funkcje dyskryminacyjne), na których podstawie różnicuje się grupy obiektów (podregionów). Najważniejszymi zagadnieniami analizy dyskryminacyjnej, które należy rozwiązać przed oszacowaniem parametrów funkcji dyskryminacyjnej, są: wybór zmiennej grupującej i wybór optymalnego zbioru cech diagnostycznych. Zadaniem wyboru zmiennej grupującej jest wstępny podział badanych obiektów na kilka grup. Może

być on dokonany np. za pomocą metody k -średnich, która należy do analizy skupień (klasyfikacji). Metoda ta polega na podziale obiektów na zadaną z góry liczbę skupień, różniącą się jak najmniej w ramach danego skupienia a jak najbardziej między różnymi skupieniami [Pociecha i in. 1998, s. 95-96].

Po zakończeniu wyżej wymienionych czynności wstępnych przystępuje się do szacowania parametrów funkcji dyskryminacyjnych a następnie do wyznaczenia ich wartości dla każdego obiektu. Wartości liniowych funkcji dyskryminacyjnych wyznaczamy według wzoru [Morrison 1990, s. 354-363]:

$$W_{ij} = x'S^{-1}(\bar{x}_i - \bar{x}_j) - \frac{1}{2}(\bar{x}_i + \bar{x}_j)'S^{-1}(\bar{x}_i - \bar{x}_j) \quad (1)$$

i postępujemy zgodnie z następującą regułą: obserwację x klasyfikujemy do i -tej grupy, jeżeli $W_{ij} > 0$ dla wszystkich $j \neq i$. Do klasyfikacji poszczególnych obiektów do jednej z trzech grup wykorzystuje się różne statystyki dyskryminacyjne:

$$W_{12} = x'S^{-1}(\bar{x}_1 - \bar{x}_2) - \frac{1}{2}(\bar{x}_1 + \bar{x}_2)'S^{-1}(\bar{x}_1 - \bar{x}_2), \quad (2)$$

$$W_{13} = x'S^{-1}(\bar{x}_1 - \bar{x}_3) - \frac{1}{2}(\bar{x}_1 + \bar{x}_3)'S^{-1}(\bar{x}_1 - \bar{x}_3), \quad (3)$$

$$W_{23} = x'S^{-1}(\bar{x}_2 - \bar{x}_3) - \frac{1}{2}(\bar{x}_2 + \bar{x}_3)'S^{-1}(\bar{x}_2 - \bar{x}_3). \quad (4)$$

Ponieważ $W_{23} = W_{13} - W_{12}$, to wystarczy zastosowanie statystyki W_{12} oraz W_{13} . Reguła klasyfikacyjna będzie mieć następującą postać:

- zaklasyfikujemy x do populacji pierwszej, jeżeli:

$$W_{12} > 0 \text{ oraz } W_{13} > W_{12} \quad (5)$$

- zaklasyfikujemy x do populacji drugiej, jeżeli:

$$W_{12} < 0 \text{ oraz } W_{13} > W_{12}, \quad (6)$$

- zaklasyfikujemy x do populacji trzeciej, jeżeli:

$$W_{13} < 0 \text{ oraz } W_{12} > W_{13}, \quad (7)$$

WYBÓR ZMIENNYCH DO BADANIA

Do klasyfikacji podregionów w Polsce wykorzystano wstępnie 18 cech (wskaźników) charakteryzujących, z jednej strony atrakcyjność środowiska naturalnego (stymulanty), a z drugiej mierzących poziom jego zanieczyszczenia (destymulanty) w 2011 roku:

X_1 – lesistość w %,

X_2 – ludność korzystająca z oczyszczalni ścieków w % ludności ogółem,

X_3 – emisja zanieczyszczeń pyłowych powietrza z zakładów szczególnie uciążliwych w tonach na 1 km²,

X_4 – emisja zanieczyszczeń gazowych powietrza z zakładów szczególnie uciążliwych w tonach na 1 km²,

- X_5 – odpady (z wyłączeniem komunalnych) wytworzone w ciągu roku w tonach na 1 km²,
 X_6 – udział odpadów składowanych w ilości odpadów wytworzonych w ciągu roku w %,
 X_7 – odpady komunalne zebrane w ciągu roku na 1 mieszkańca w kg,
 X_8 – ścieki przemysłowe i komunalne oczyszczane w % ścieków wymagających oczyszczania,
 X_9 – ścieki komunalne oczyszczane na 100 km² w dam³,
 X_{10} – powierzchnia o szczególnych walorach przyrodniczych prawnie chroniona w % powierzchni ogółem,
 X_{11} – powierzchnia parków narodowych w % powierzchni ogółem,
 X_{12} – powierzchnia rezerwatów przyrody w % powierzchni ogółem,
 X_{13} – powierzchnia parków krajobrazowych w % powierzchni ogółem,
 X_{14} – obszary chronionego krajobrazu w % powierzchni ogółem,
 X_{15} – powierzchnia użytków ekologicznych w % powierzchni ogółem,
 X_{16} – pomniki przyrody na 1 km²,
 X_{17} – nakłady na środki trwałe służące gospodarce wodnej w tys. zł. na 1 mieszkańca,
 X_{18} – nakłady na środki trwałe służące ochronie środowiska w tys. zł. na 1 mieszkańca.

Po określeniu i zgromadzeniu danych dotyczących wstępnego zestawu cech podejmuje się zazwyczaj w odniesieniu do nich odpowiednie działania weryfikacyjne według dwóch najistotniejszych kryteriów [Młodak 2006, s. 28-32]:

1. Zmienność – cechy powinny wykazywać odpowiednie zróżnicowanie, czyli skutecznie dyskryminować obiekty. Do oceny zmienności służy współczynnik zmienności obliczany według wzoru:

$$V_j = \frac{S_j}{\bar{x}_j}, \quad (8)$$

gdzie: \bar{x}_j - średnia arytmetyczna wartości cechy X_j , S_j - odchylenie standardowe j -tej cechy, $j = 1, 2, \dots, m$, m - liczba cech.

2. Korelacja – dwie cechy silnie ze sobą skorelowane są nośnikami podobnej informacji, a więc jedna z nich staje się zbędna. Dlatego należy wziąć pod uwagę współczynniki korelacji wszystkich par cech, a następnie zastosować odpowiednią metodę weryfikacji w celu wyeliminowania cech najbardziej podobnych do innych. Punktem wyjścia jest wyznaczenie macierzy korelacji cech:

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & 1 \end{bmatrix} \quad (9)$$

gdzie: r_{jk} - współczynnik korelacji liniowej Persony j -tej i k -tej cechy.

Biorąc pod uwagę pierwsze z wyżej wymienionych kryteriów z badania wyeliminowano tylko zmienną X_8 , dla której współczynnik zmienności nie przekraczał 10%.

W kolejnym kroku wyznaczono macierz korelacji pomiędzy cechami. Do dyskryminacji zmiennych zastosowano metodę odwróconej macierzy [Panek 2009, s. 22-23]. Polega ona na wyznaczeniu macierzy odwrotnej do macierzy \mathbf{R} , czyli:

$$\mathbf{R}^{-1} = \begin{bmatrix} \tilde{r}_{11} & \tilde{r}_{12} & \dots & \tilde{r}_{1m} \\ \tilde{r}_{21} & \tilde{r}_{22} & \dots & \tilde{r}_{2m} \\ \dots & \dots & \dots & \dots \\ \tilde{r}_{m1} & \tilde{r}_{m2} & \dots & \tilde{r}_{mm} \end{bmatrix}, \quad (10)$$

w której

$$\tilde{r}_{jk} = \frac{(-1)^{j+k} \det(\mathbf{R}_{kj})}{\det(\mathbf{R})}, \quad (11)$$

przy czym $\det(\mathbf{R})$ – wyznacznik macierzy \mathbf{R} , \mathbf{R}_{kj} – oznacza macierz powstałą z macierzy po usunięciu z niej j -tego wiersza i k -tej kolumny ($j, k = 1, 2, \dots, m$).

Elementy diagonalne macierzy \mathbf{R}^{-1} przyjmują wartości z przedziału $[1, \infty)$. Te z nich, które przekraczają ustalony maksymalny poziom \tilde{r}_0 (często przyjmuje się $\tilde{r}_0 = 10$) świadczą o wadliwym uwarunkowaniu numerycznym macierzy \mathbf{R} . Należy, więc dokonać eliminacji tych cech, dla których $|\tilde{r}_{jj}| > \tilde{r}_0$.

Wykorzystanie metody odwróconej macierzy współczynników korelacji doprowadziło do uzyskania następującego zbioru cech: $X_2, X_3, X_5, X_7, X_9, X_{16}$. Wszystkie przyjęte do badania cechy, poza X_{16} to destymulanty.

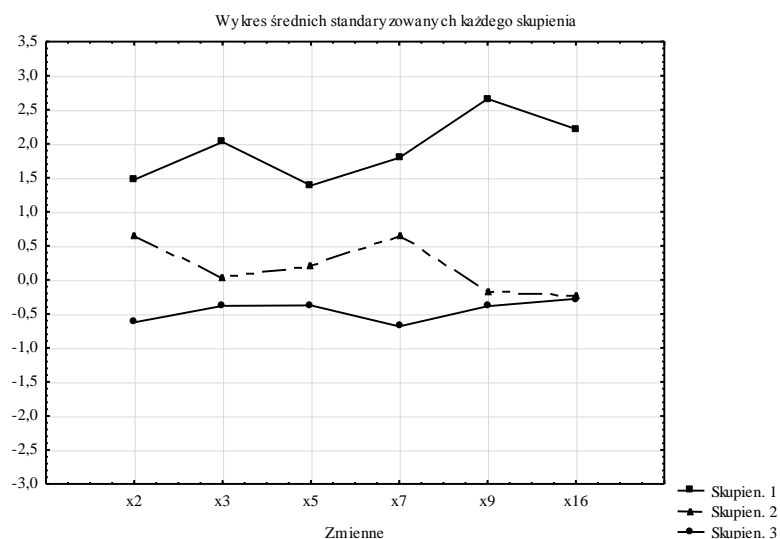
WYKORZYSTANIE ANALIZY DYSKRYMINACYJNEJ DO KLASYFIKACJI PODREGIONÓW W POLSCE

Wstępnej klasyfikacji obiektów (podregionów) na trzy grupy, a tym samym wyboru zmiennej grupującej dokonano stosując metodę k -średnich. Podstawą klasyfikacji były standaryzowane wartości cech diagnostycznych przeprowadzone według formuły [Nowak 1990, s. 38]:

$$z_{jk} = \frac{x_{jk} - \bar{x}_j}{S_j}. \quad (12)$$

Standaryzacja pozwoliła wyeliminować wpływ jednostek, w których te cechy są wyrażone, na wynik grupowania. We wstępnym podziale do najgorszej pierwszej grupy ze względu na badane cechy diagnostyczne zaliczono siedem podregionów, trzydzieści siedem zostało zaliczonych do grupy najlepszej, a pozostałe dwadzieścia dwa zaliczono do grupy pośredniej. Wybrane zmienne dobrze dyskryminują badane obiekty na trzy skupienia, co potwierdza wykres średnich wartości zmiennych standaryzowanych dla poszczególnych skupień (Rys. 1).

Rysunek 1. Wartości średnie standaryzowane zmiennych diagnostycznych w poszczególnych skupieniach.



Źródło: obliczenia własne

Na podstawie przyjętych cech diagnostycznych oszacowano funkcje dyskryminacyjne. O statystycznie istotnej dyskryminacji podregionów świadczy bliska zero statystyka λ Wilksa ($\lambda = 0,041$). Na podstawie wartości funkcji dyskryminacyjnych, korzystając z reguł klasyfikacyjnych danych wzorami (5) – (7) poszczególne obiekty (podregiony) zaliczono do odpowiednich grup. W celu określenia jakości klasyfikacji wyznaczono prawdopodobieństwa *a priori* zaliczenia poszczególnych obiektów do jednej z trzech grup przy danym z góry prawdopodobieństwie *a posteriori*. Obiekt zaliczany jest do tej grupy, dla której prawdopodobieństwo *a priori* przyjmuje wartość maksymalną [Zawadzki 1999, s. 12-13]. W badaniu przyjęto, że prawdopodobieństwa *a posteriori* są proporcjonalne do liczebności skupień i wynoszą odpowiednio: 11, 33 i 56 procent. W tabeli 1 zamieszczone zostały prawdopodobieństwa *a priori* i przynależności

podregionów do poszczególnych grup oraz wartości zmiennej grupującej wyznaczone metodą k -średnich, czyli w klasyfikacji wstępnej.

Tabela 1. Klasyfikacja podregionów na grupy i prawdopodobieństwo a priori przynależności do grupy

Podregiony	Klasyfikacja wstępna	Prawdopodobieństwo <i>a priori</i> przynależności do grupy		
		1	2	3
m. Łódź	G_1:1	1,00	0,00	0,00
m. Warszawa	G_1:1	1,00	0,00	0,00
m. Kraków	G_1:1	1,00	0,00	0,00
katowicki	G_1:1	1,00	0,00	0,00
m. Poznań	G_1:1	1,00	0,00	0,00
m. Wrocław	G_1:1	1,00	0,00	0,00
trójmiejski	G_1:1	1,00	0,00	0,00
łódzki	G_2:2	0,00	0,88	0,12
warszawski zachodni	G_2:2	0,00	0,81	0,19
bytomski	G_2:2	0,00	1,00	0,00
gliwicki	G_2:2	0,00	1,00	0,00
rybnicki	G_2:2	0,00	0,91	0,09
sosnowiecki	G_2:2	0,00	1,00	0,00
tyski	G_2:2	0,00	1,00	0,00
białostocki	G_2:2	0,00	79,00	0,21
gorzowski	G_2:2	0,00	0,99	0,01
zielonogórski	G_2:2	0,00	0,99	0,01
koszaliński	G_2:2	0,00	0,98	0,02
stargardzki	G_2:2	0,00	0,77	0,23
m. Szczecin	G_2:2	0,01	0,99	0,00
szczeciński	G_2:2	0,00	0,96	0,04
jeleniogórski	G_2:2	0,00	0,99	0,01
legnicko-głogowski	G_2:2	0,00	1,00	0,00
wałbrzyski	G_2:2	0,00	0,99	0,01
opolski	G_2:2	0,00	0,92	0,08
bydgosko-toruński	G_2:2	0,00	1,00	0,00
śląski	G_2:2	0,00	0,95	0,05
starogardzki	G_2:2	0,00	0,80	0,20
olsztyński	G_2:2	0,00	0,75	0,25
piotrkowski	G_3:3	0,00	0,00	1,00
sieradzki	G_3:3	0,00	0,00	1,00
skierniewicki	G_3:3	0,00	0,00	1,00
ciechanowsko-płocki	G_3:3	0,00	0,00	1,00
ostrołęcko-siedlecki	G_3:3	0,00	0,00	1,00
radomski	G_3:3	0,00	0,00	1,00
warszawski wschodni	G_3:3	0,00	0,01	0,99
krakowski	G_3:3	0,00	0,00	1,00

Podregiony	Klasyfikacja wstępna	Prawdopodobieństwo <i>a priori</i> przynależności do grupy		
		1	2	3
nowosądecki	G_3:3	0,00	0,00	1,00
oświęcimski	G_3:3	0,00	0,00	1,00
tarnowski	G_3:3	0,00	0,00	1,00
bielski	G_3:3	0,00	0,25	0,75
częstochowski	G_3:3	0,00	0,02	0,98
białski	G_3:3	0,00	0,00	1,00
chełmsko-zamojski	G_3:3	0,00	0,00	1,00
lubelski	G_3:3	0,00	0,00	1,00
puławski	G_3:3	0,00	0,00	1,00
krośnieński	G_3:3	0,00	0,00	1,00
przemyski	G_3:3	0,00	0,03	0,97
rzeszowski	G_3:3	0,00	0,01	0,99
tarnobrzegi	G_3:3	0,00	0,01	0,99
łomżyński	G_3:3	0,00	0,00	1,00
suwalski	G_3:3	0,00	0,07	0,93
kielecki	G_3:3	0,00	0,04	0,96
sandomiersko-jędrzejowski	G_3:3	0,00	0,00	1,00
kałiski	G_3:3	0,00	0,01	0,99
koniński	G_3:3	0,00	0,00	1,00
leszczyński	G_3:3	0,00	0,05	0,95
piłski	G_3:3	0,00	0,14	0,86
poznański	G_3:3	0,00	0,18	0,82
wrocławski	G_3:3	0,00	0,01	0,99
nyski	G_3:3	0,00	0,02	0,98
grudziądzki	G_3:3	0,00	0,01	0,99
włocławski	G_3:3	0,00	0,00	1,00
gdański	G_3:3	0,00	0,14	0,86
elbląski	G_3:3	0,00	0,08	0,92
ęlecki	G_3:3	0,00	0,03	0,97

Źródło: obliczenia własne

Z informacji zawartych w tabeli wynika, że prawie wszystkie prawdopodobieństwa są równe lub zbliżone do jedności. Oznacza to, że dla przyjętego zestawu cech diagnostycznych przydział obiektów do poszczególnych skupień jest jednoznaczny. Świadczy to o identyczności klasyfikacji wstępnej wyznaczonej metodą k-średnich i ostatecznej, otrzymanej na podstawie funkcji dyskryminacyjnych.

Z kolumny drugiej tabeli wynika, że w najgorszej pierwszej grupie znalazło się siedem podregionów, wśród których dominują duże miasta oraz ośrodki przemysłowe. Z uwagi na intensywnie prowadzoną działalność gospodarczą i duże skupiska ludności charakteryzują się one wysokim stopniem zanieczyszczenia

środowiska naturalnego. Nie sprzyja to z całą pewnością rozwojowi turystyki przyjaznej środowisku, mimo, że w podregionach tych liczba pomników przyrody kształtuje się powyżej przeciętnej w kraju. Grupa trzecia obejmuje aż 37 podregionów o najbardziej korzystnych wartościach wybranych cech. Charakteryzują się one przede wszystkim stosunkowo niskim stopniem zanieczyszczenia środowiska naturalnego (mała ilość odpadów przemysłowych i komunalnych, niska emisja zanieczyszczeń pyłowych powietrza). Do grupy drugiej zakwalifikowano 22 obiekty, charakteryzujące się dużym zróżnicowaniem wartości cech diagnostycznych w poszczególnych podregionach.

PODSUMOWANIE

Z przeprowadzonych badań wynika, że analiza dyskryminacyjna może być z powodzeniem wykorzystywana w wielowymiarowej klasyfikacji obiektów. Jej stosowanie wydaje się zasadne zwłaszcza z uwagi na prostotę budowy, łatwość interpretacji i ogólną dostępność procedury w pakietach statystycznych.

Podregiony w Polsce wykazują przestrzenne zróżnicowanie poziomu atrakcyjności z punktu widzenia możliwości rozwoju w nich turystyki przyjaznej środowisku naturalnemu. Obecnie turystyka odgrywa coraz ważniejszą rolę w rozwoju jednostek terytorialnych. Stanowi ona ważny czynnik podnoszący atrakcyjność i konkurencyjność regionu. Dlatego ważne jest prawidłowe i kompleksowe zdiagnozowanie uwarunkowań rozwoju funkcji turystycznej obszaru recepcji. Formą turystyki, która cieszy się coraz większym zainteresowaniem jest właśnie ekoturystyka. Może być ona szansą rozwoju obszarów szczególnie cennych przyrodniczo. Współczesna rola gospodarki turystycznej polega na takim korzystaniu z zasobów materialnych i niematerialnych środowiska, aby była zachowana równowaga w sferze nie tylko ekologicznej, ale też społecznej, ekonomicznej i przestrzennej. Najważniejszym zaś skutkiem rozwoju turystyki przyjaznej dla środowiska jest wzrost poziomu i jakości życia ludzi, zarówno mieszkańców regionów recepcji turystycznej, jak i samych turystów.

Wzięcie pod uwagę ekologicznej charakterystyki regionów powinno być podstawą prowadzenia właściwej polityki gospodarczej, w tym inwestycyjnej, zarówno przez państwo, jak i przez samorządy. Ekoturystyka, powiązana z pokrewnymi jej formami turystyki zrównoważonej, drobną przedsiębiorczością, rolnictwem, przetwórstwem i tradycyjnym rzemiosłem, może być wyzwaniem dla przeżywających kryzys społeczno-ekonomiczny terenów objętych bezrobociem, może stać się dziedziną stymulującą rozwój tych rejonów.

BIBLIOGRAFIA

- Karczmarska A. (2010) Możliwości rozwoju agro- i ekoturystyki w Polsce na przykładzie województwa śląskiego [w:] Acta Scientiarum Poloniarum Oeconomia 9(4), Wydawnictwo SGGW, Warszawa.
- Łobożewicz T., Bieńczyk G. (2001) Podstawy turystyki, Wyższa Szkoła Ekonomiczna, Warszawa.
- Mika M. (2008) Charakterystyka wybranych form turystyki [w:] Turystyka, red. W. Kurek, Wydawnictwo Naukowe PWN, Warszawa.
- Młodak A. (2006) Analiza taksonomiczna w statystyce regionalnej, Difin, Warszawa.
- Morrison D. (1990) Wielowymiarowa analiza statystyczna, PWN, Warszawa.
- Panek T., (2009) Statystyczne metody wielowymiarowej analizy porównawczej, Warszawa, Oficyna Wydawnicza SGH.
- Pociecha J., Podolec B., Sokołowski A., Zając K., (1998) Metody taksonomiczne w badaniach społeczno-ekonomicznych, Państwowe Wydawnictwo Naukowe, Warszawa.
- Zaręba D. (2000) Ekoturystyka. Wyzwania i nadzieje, Wydawnictwo Naukowe PWN, Warszawa.
- Zawadzki J. (1999) Zastosowanie analizy dyskryminacyjnej dla wielu populacji do badania kondycji finansowej firm, Instytut Ekonomii i Zarządzania Politechniki Szczecińskiej, Wydawnictwo Zapol, Szczecin.

APPLICATION OF DISCRIMINANT ANALYSIS IN THE STUDY OF LEVEL OF DIVERSITY OF ECOTOURISM IN POLAND

Abstract: The main goal of this paper is the analysis of the spatial differentiation of Poland's sub-regions in terms of the attractiveness, regarding the development of the environment friendly tourism, the so called ecotourism. In the analysis were used factors indicating the attractiveness of natural environment in Poland's sub-regions (stimulants) and factors that measure the pollution (destimulants). The classification of the sub-regions was performed with discrimination analysis. The initial classification of objects into groups, thereby selecting the grouping variable, was made using the *k*-means method.

Keywords: ecotourism, discrimination analysis, *k*-means method

DYNAMIKA REGIONALNEGO ZRÓŻNICOWANIA ROLNICTWA W POLSCE W LATACH 1998-2010

Agata Binderman

Katedra Ekonometrii i Statystyki
Szkola Główna Gospodarstwa Wiejskiego w Warszawie
e-mail: agata_binderman@sggw.pl

Streszczenie: Praca jest bezpośrednią kontynuacją badań autorki, których wyniki zamieszczone zostały w publikacji [Binderman 2012]. W niniejszej pracy korzystając z uzyskanych wyników dokonano analiz w zakresie poziomu oraz zróżnicowania rolnictwa polskich województw w latach 1998-2010. Przeprowadzone badania pozwoliły na zmierzenie i ocenę dynamiki rozwoju poziomu rolnictwa w Polsce oraz sporządzenie klasyfikacji województw ze względu na zmiany w badanym okresie.

Słowa kluczowe: miernik syntetyczny, poziom rozwoju rolnictwa, klasyfikacja, dynamika zmian

WSTĘP

Zjawiska gospodarczo-społeczne w rolnictwie, szczególnie w ostatnich latach charakteryzują się tym, że ulegają bardzo szybkim zmianom. Wejście Polski do Unii Europejskiej, intensyfikacja procesu integracji europejskiej w ramach Wspólnej Polityki Rolnej zdyktowało zmiany na wsi. Wprowadzenie polityki rozwoju obszarów wiejskich przyspieszyło proces modernizacji sektora rolnego i dywersyfikacji źródeł dochodu na wsi. [Urząd Komitetu Integracji Europejskiej 2009, Binderman 2007]. Wejście Polski do UE zaowocowało wzrostem poziomu dochodów gospodarstw rolnych, co wpłynęło na wzrost inwestycji na wsi i przyspieszyło procesy modernizacyjne gospodarstw. Członkostwo Polski w Unii Europejskiej spowodowało wzrost poziomu eksportu do państw unijnych, pozytywnie wpłynęło na działalność produkcyjną polskiego rolnictwa. Poprawa rentowności sektora zachęciła rolników do rozwoju i intensyfikacji produkcji. Nastąpił wzrost areалу zasiewów, który był wynikiem uruchomienia płatności bezpośrednich oraz płatności dla obszarów o niekorzystnych warunkach

gospodarowania (ONW). Wprowadzenie Wspólnej Polityki Rolnej uruchomiło instrumenty służące stabilizacji rynków, podtrzymywało ceny produktów rolnych, sprzyjało to opłacalności produkcji, mimo dużego wzrost cen środków produkcji.

Z powyższych powodów istnieje konieczność analizy tych zjawisk w czasie, czyli ich dynamiki, analizy różnymi narzędziami ekonometryczno – statystycznymi. Problematyka regionalnego zróżnicowania rolnictwa w Polsce była przez autorkę badana w kilkunastu pracach [Binderman 2004, 2005 a, b, 2006 a, b, 2007, 2008 a, b, 2009, 2010, 2011, 2012]. Z licznej literatury związanej z rozważaną problematyką, wymienimy [Harasim (red) 2006, Krasowicz 2008, 2009, Muszyńska 2010].

W pracy [Binderman 2012] przy pomocy wielowymiarowej analizy porównawczej dokonano oceny regionalnego zróżnicowania rolnictwa w Polsce, w latach 1998-2010 - bez analizy dynamiki zjawisk. W rozważanym okresie, przy pomocy mierników różnych syntetycznych, mających charakter funkcji użyteczności dokonano klasyfikacji i uporządkowania województw. Na podstawie wyników, otrzymanych w cytowanej pracy, można dokonać analizy dynamiki, która będzie miała na celu określenie rozmiarów i kierunków rozwoju, czyli zmian w czasie badanego zjawiska. Do ustalenia stopnia wzrostu lub spadku poziomu rolnictwa dla poszczególnych województw oraz tendencji rozwojowych zastosowano w pracy podstawowe indeksy, spełniające rolę wskaźników dynamiki [Michalski 2004].

METODYKA BADAŃ

W pracy [Binderman 2012] autorka zbadała zróżnicowanie poziomu rozwoju rolnictwa, przyjmując jeden sposób normalizacji zmiennych, jeden wzór określający odległość między rozważanymi obiektami i cztery różne mierniki syntetyczne tych obiektów – wzorcowych jak i bezwzorcowego. Po normalizacji zmiennych przy pomocy unitaryzacji zerowanej [Kukuła 2000] w trzech pierwszych miernikach wykorzystano metodę bezwzorcową MUZ [Cieślak 1993, Kukuła 2000], metodę TOPSIS [Hwang, Yoon 1981] oraz metodę podaną przez autorkę [Binderman 2006a, 2007a, 2011], czwarty miernik był średnią geometryczną trzech poprzednich.

Użyty miernik bezwzorcowy dla wektora opisanego przez znormalizowane cechy w danym roku, danego województwa ($13 \times 16 = 208$), został określony za pomocą wzoru:

$$m_1(\mathbf{z}_i) = \frac{1}{10} \sum_{j=1}^{10} z_{ij}; i = 1, \dots, 208 \quad (1)$$

W metodach wzorcowych wykorzystano dwa wzorce $\mathbf{z}_0 = \mathbf{0}$, $\mathbf{z}_{209} = \mathbf{1}$, które reprezentują „najgorsze”, „najlepsze” hipotetyczne województwo w

rozważanym okresie, odpowiednio. Mierniki wzorcowe określone są dla danego wektora $\mathbf{z}_i \in [\mathbf{0}, \mathbf{1}]$, $i = 0, 1, \dots, 209$, za pomocą wzorów:

$$m_2(\mathbf{z}_i) = \frac{d(\mathbf{0}, \mathbf{z}_i)}{d(\mathbf{0}, \mathbf{z}_i) + d(\mathbf{1}, \mathbf{z}_i)}, \quad (2)$$

$$m_3(\mathbf{z}_i) = \frac{1}{2} + \frac{d(\mathbf{0}, \mathbf{z}_i) - d(\mathbf{1}, \mathbf{z}_i)}{2d(\mathbf{0}, \mathbf{1})} \quad (3)$$

gdzie $d(\mathbf{x}, \mathbf{y}) = \left(\sum_{j=1}^n (x_j - y_j)^4 \right)^{\frac{1}{4}}$ dla $\mathbf{x} = (x_1, x_2, \dots, x_n)$, $\mathbf{y} = (y_1, y_2, \dots, y_n) \in \mathcal{R}_+^n$.

Miernik będący funkcją trzech poprzednich mierników został określony za pomocą wzoru:

$$m(\mathbf{z}_i) = \sqrt[3]{m_1(\mathbf{z}_i) m_2(\mathbf{z}_i) m_3(\mathbf{z}_i)}, \quad \mathbf{z}_i \in [\mathbf{0}, \mathbf{1}], \quad i = 0, 1, \dots, 209. \quad (4)$$

Można pokazać, że podane wyżej mierniki są znormalizowane tj.

$$0 \leq m_k(\mathbf{z}_i) \leq 1 \quad \text{dla} \quad \mathbf{z}_i \in [\mathbf{0}, \mathbf{1}], \quad k = 1, 2, 3, 4.$$

Metody te pozwoliły otrzymać cztery różne macierze \mathbf{M}^1 , \mathbf{M}^2 , \mathbf{M}^3 , \mathbf{M} opisujące wartości mierników syntetycznych danego województwach w danym roku, okresu 1998 - 2010, względem każdego z zastosowanych mierników, odpowiednio. Każda z tych macierzy posiada szesnaście wierszy, które odpowiadają województwom i trzynaście kolumn, które odpowiadają kolejnym latom okresu 1998 - 2010.

Na podstawie wymienionych wyżej macierzy dokonano analizy dynamiki, która miała na celu określenie rozmiarów i kierunków rozwoju, czyli zmian w czasie badanego zjawiska. Do ustalenia stopnia wzrostu lub spadku poziomu rolnictwa dla poszczególnych województw oraz tendencji rozwojowych zastosowano następujące indeksy, spełniające rolę wskaźników dynamiki [Michalski 2004]. Na przykład, dla macierzy $\mathbf{M} = [m_{ik}]_{16 \times 13}$ będącej wynikiem obliczeń przeprowadzonych według wzoru (4) miary dynamiki przedstawiają poniższe wzory.

Przyrost względny danego województwa w poszczególnych latach (w stosunku do roku poprzedniego):

$$I_{ik}^{pw} = \frac{m_{ik+1} - m_{ik}}{m_{ik}}, \quad i = 1, 2, \dots, 16; \quad k = 2, 3, \dots, 13.$$

Tempo wzrostu (w %): $T_{ik}^w = I_{ik}^{pw} \times 100$.

Indeks łańcuchowy w % (rok poprzedni = 100): $I_{ik}^l = \frac{m_{ik+1}}{m_{ik}} \times 100\%$.

Indeks ułamkowy (o podstawie stałej): $I_{ik}^u = \frac{m_{ik}}{m_{i1}}$.

Indeks procentowy (o podstawie stałej): $I_{ik}^p = I_{ik}^u \times 100$.

Indeks o podstawie zmiennej (łańcuchowy) w %:

$$I_{ik+1}^z = \frac{m_{ik+1}}{m_{ik}} \times 100, \quad i = 1, \dots, 16; \quad k = 1, 2, \dots, 13.$$

Średnie tempo zmian danego województwa:

$$G_i = \sqrt[12]{I_{i2}^z \cdot I_{i3}^z \cdot \dots \cdot I_{i13}^z}, \quad i = 1, \dots, 16.$$

WYNIKI BADAŃ

Niech $M^j = [m_{ik}^j]_{16 \times 13}$, $j = 1, 2, 3$, $M = [m_{ik}]_{16 \times 13}$ oznaczają macierz utworzoną z mierników poziomu rozwoju rolnictwa poszczególnych województw, w poszczególnych latach, tj. m_{ik} – wartość miernika i -tego województwa (dolnośląskie – 1, ..., zachodniopomorskie – 16) w k -tym roku (1998 - 1, 1999 - 2, ..., 2010 - 13), według wzorów (1) - (4), odpowiednio. Poniżej podane zostały elementy rozważanych macierzy M^1 , M^2 , M^3 i M :

$$M = [m_{ik}]_{16 \times 13} =$$

0,38	0,38	0,37	0,36	0,38	0,34	0,38	0,37	0,34	0,39	0,40	0,43	0,43
0,42	0,41	0,38	0,40	0,39	0,38	0,41	0,39	0,38	0,41	0,44	0,48	0,48
0,46	0,42	0,43	0,44	0,42	0,39	0,41	0,40	0,34	0,39	0,42	0,49	0,47
0,23	0,25	0,23	0,24	0,22	0,19	0,28	0,24	0,22	0,27	0,27	0,30	0,34
0,38	0,37	0,34	0,38	0,35	0,33	0,35	0,35	0,33	0,37	0,39	0,46	0,47
0,33	0,32	0,32	0,32	0,33	0,31	0,31	0,30	0,29	0,33	0,33	0,37	0,39
0,40	0,39	0,37	0,39	0,37	0,35	0,38	0,39	0,37	0,42	0,46	0,53	0,51
0,45	0,43	0,44	0,44	0,43	0,41	0,45	0,44	0,42	0,47	0,49	0,49	0,51
0,29	0,29	0,28	0,29	0,29	0,29	0,28	0,26	0,25	0,28	0,28	0,31	0,31
0,36	0,37	0,36	0,37	0,36	0,36	0,36	0,38	0,34	0,39	0,42	0,43	0,45
0,34	0,34	0,32	0,32	0,33	0,31	0,33	0,33	0,31	0,35	0,38	0,41	0,40
0,27	0,26	0,24	0,25	0,25	0,24	0,27	0,25	0,24	0,32	0,36	0,38	0,36
0,35	0,35	0,32	0,35	0,33	0,31	0,33	0,31	0,27	0,30	0,32	0,38	0,38
0,36	0,36	0,34	0,35	0,35	0,37	0,36	0,36	0,35	0,39	0,41	0,43	0,43
0,40	0,39	0,38	0,41	0,41	0,38	0,43	0,41	0,39	0,46	0,51	0,57	0,52
0,36	0,37	0,35	0,34	0,35	0,35	0,37	0,34	0,35	0,37	0,37	0,39	0,42

Tabela 1 przedstawia indeksy procentowe województw I_{ik}^p (gdzie w_1 to województwo dolnośląskie, w_2 kujawsko-pomorskie, itd., a w ostatnim wierszu dla Polski) o podstawie stałej dla poszczególnych lat badanego okresu (1998 = 100).

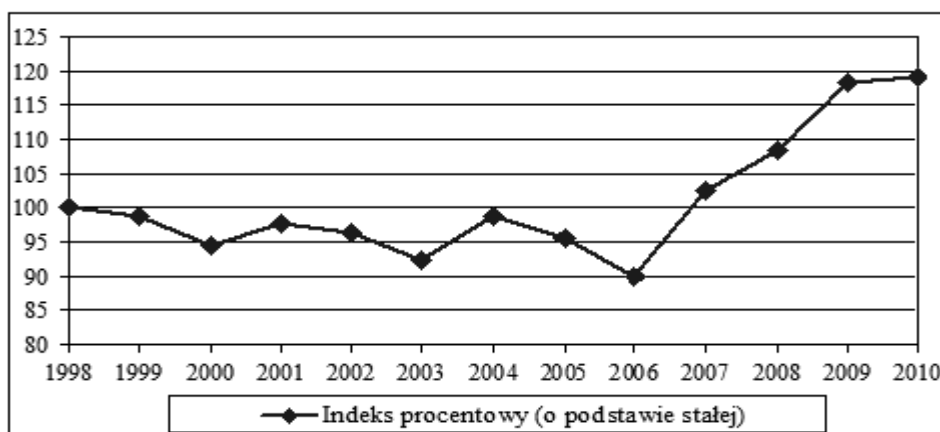
Tabela 1. Indeksy procentowe województw w latach 1998-2010

	'99	'00	'01	'02	'03	'04	'05	'06	'07	'08	'09	'10
w ₁	99,7	96,4	95,4	98,4	90,3	100,5	96,1	89,9	102,1	104,1	111,9	113,3
w ₂	98,8	90,5	96,5	94,2	90,5	97,2	93,1	89,6	99,1	104,2	114,1	115,0
w ₃	92,0	92,4	94,5	90,7	84,7	89,0	87,4	74,1	85,0	91,8	107,1	102,9
w ₄	109,2	100,8	103,7	97,2	83,1	121,0	102,7	96,9	115,8	117,0	133,1	149,7
w ₅	99,0	91,3	100,9	93,3	87,4	93,6	93,2	88,9	99,4	104,7	122,6	124,3
w ₆	98,8	96,7	97,3	99,8	93,2	95,4	91,4	87,1	100,4	101,5	111,4	117,5
w ₇	98,4	92,7	97,9	92,4	89,0	95,4	97,0	93,0	106,4	116,0	133,7	128,8
w ₈	94,3	97,5	96,3	96,2	91,5	100,5	98,3	92,7	104,9	109,1	108,0	113,7
w ₉	99,0	96,6	98,1	100,7	100,8	94,5	88,6	86,4	94,6	96,8	106,1	105,6
w ₁₀	101,5	99,6	101,5	101,1	99,4	99,0	105,5	95,2	106,8	116,0	118,6	125,0
w ₁₁	99,2	93,6	93,9	97,6	90,5	95,5	96,9	91,2	102,9	112,2	119,6	117,1
w ₁₂	95,4	87,6	94,3	91,1	89,9	101,8	94,1	88,8	118,0	134,5	140,3	133,9
w ₁₃	99,2	92,2	101,0	93,4	89,4	95,9	88,1	78,1	87,1	91,1	107,8	108,6
w ₁₄	99,4	93,4	96,8	97,9	103,5	100,0	99,9	97,6	108,7	114,3	119,1	120,2
w ₁₅	97,5	94,7	100,7	100,3	94,7	105,7	101,5	96,9	113,9	125,3	140,3	129,6
w ₁₆	104,5	97,2	97,0	97,7	98,2	103,8	97,0	97,7	105,4	103,1	111,1	118,7
P	98,7	94,5	97,7	96,2	92,2	98,7	95,6	90,0	102,5	108,3	118,4	119,2

Źródło: obliczenia własne

Poniższy rysunek 1 przedstawia wykres indeksu procentowego o podstawie stałej dla Polski w badanym okresie (1998 = 100).

Rysunek 1. Indeks procentowy (o podstawie stałej) dla Polski w latach 1998-2010



Źródło: opracowanie własne

W dynamicznym ujęciu zmian poziomu polskiego rolnictwa, podstawą stała była wartość obliczona dla 1998 roku. To umożliwiło porównanie przemian, które miały miejsce w całym obserwowanym okresie, w stosunku do sytuacji w Polsce, w roku 1998. Wyniki otrzymane z tak przeprowadzonych badań pokazują dynamikę zmian (zarówno wzrostową jak i spadkową) syntetycznych mierników, które opisują poziom rolnictwa Polski w latach 1999–2010, w stosunku do wartości tych mierników na początku rozważanego okresu. Celem analizowania dynamiki zmian wartości syntetycznych mierników dla poszczególnych województw i dla Polski było wskazanie, w których województwach poziom rozwoju rolnictwa w danym czasie zwiększył się, w którym obniżył lub pozostał stały.

Wyniki indeksów procentowych o podstawie stałej (rok 1998 = 100) pokazują, że do roku 2006 występowały wahania poziomu rozwoju rolnictwa Polski - spadki w latach 1999 - 2000, 2002 - 2003 oraz 2005 - 2006 (por. rys. 1), natomiast w pozostałych latach tj. w 2001, 2004 i od 2007 do 2010 występował wzrost indeksu. W latach 1999 - 2005 dynamika zmian większości województw w stosunku do roku 1998 była ujemna (por. tabela 1), a w roku 2006 we wszystkich województwach były spadki indeksów. Natomiast w kolejnych latach tj. 2007 i 2008 w porównaniu z rokiem 1998, w większości województw zanotowano wzrost. W ostatnich dwóch latach badanego okresu wszystkie województwa miały znaczące dodatnie tempo wzrostu w stosunku do roku 1998. W ostatnim roku tj. 2010 województwo lubelskie było regionem, w którym ten wzrost był najniższy (3%), natomiast w lubuskim - najwyższy (50%).

Tabela 2 przedstawia indeksy o zmiennej podstawie I_{ik+1}^z (w %) w poszczególnych latach 1998 - 2010 dla województw (gdzie numer wiersza to numer województwa w kolejności alfabetycznej tj. dolnośląskie - 1, kujawsko-pomorskie - 2, itd.) i w ostatnim wierszu dla Polski, natomiast Rysunek 2 przedstawia tempo wzrostu T_{ik}^w (w %) dla Polski w badanym okresie.

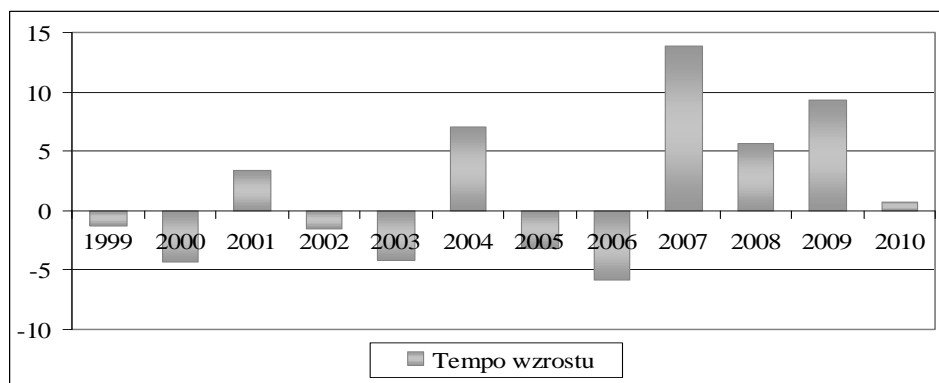
Tabela 2. Indeksy łańcuchowe o zmiennej podstawie (w %) w latach 1998-2010

	'99	'00	'01	'02	'03	'04	'05	'06	'07	'08	'09	'10
1	99,7	96,7	98,9	103,2	91,7	111,2	95,7	93,6	113,6	101,9	107,5	101,3
2	98,8	91,6	106,6	97,6	96,1	107,4	95,8	96,3	110,5	105,2	109,5	100,9
3	92,0	100,3	102,3	96,0	93,4	105,0	98,3	84,8	114,7	108,0	116,7	96,0
4	109,2	92,3	102,9	93,7	85,4	145,7	84,8	94,4	119,6	101,0	113,8	112,5
5	99,0	92,2	110,5	92,5	93,7	107,0	99,6	95,4	111,7	105,4	117,1	101,4
6	98,8	97,9	100,6	102,6	93,3	102,3	95,8	95,3	115,2	101,1	109,8	105,4
7	98,4	94,2	105,6	94,3	96,3	107,3	101,7	95,8	114,4	109,0	115,3	96,3
8	94,3	103,3	98,7	99,9	95,1	109,9	97,8	94,4	113,1	104,0	99,0	105,3
9	99,0	97,6	101,5	102,7	100,1	93,7	93,8	97,5	109,5	102,3	109,6	99,5
10	101,5	98,1	101,9	99,6	98,3	99,6	106,6	90,2	112,2	108,6	102,3	105,4
11	99,2	94,3	100,4	103,9	92,7	105,5	101,5	94,2	112,8	109,1	106,6	97,9
12	95,4	91,8	107,7	96,6	98,7	113,3	92,4	94,4	132,9	114,0	104,2	95,4

	'99	'00	'01	'02	'03	'04	'05	'06	'07	'08	'09	'10
13	99,2	92,9	109,6	92,5	95,7	107,3	91,9	88,7	111,5	104,6	118,3	100,8
14	99,4	94,1	103,6	101,2	105,7	96,6	99,9	97,7	111,3	105,1	104,2	100,9
15	97,5	97,1	106,3	99,6	94,4	111,6	96,1	95,4	117,6	110,0	112,0	92,4
16	104,5	93,0	99,8	100,6	100,5	105,7	93,5	100,7	107,9	97,8	107,8	106,8
P	98,7	95,7	103,4	98,5	95,8	107,0	96,9	94,2	113,9	105,6	109,4	100,7

Źródło: obliczenia własne

Rysunek 2. Tempo wzrostu T_{ik}^w (w %) dla Polski w latach 1998-2010



Źródło: opracowanie własne

Analiza tempa wzrostu przy zmiennej podstawie, gdzie mierniki syntetyczne w danym roku są przyrównywane do roku poprzedniego wskazuje, że w latach 1999 - 2000, 2002 - 2003 oraz 2005 - 2006 w Polsce zanotowano spadki (por. tabela 2 i rysunek 2). Największe spadki w stosunku do roku poprzedniego wystąpiły w 1999 r. w województwie lubelskim - 8%, a w 2000 r. w kujawsko-pomorskim - 8,4%. W roku 2001 w stosunku do roku 2000, znaczna większość województw charakteryzowała się wzrostem poziomu rolnictwa. Największy przyrost w 2001 roku tj. 10,5%, wystąpił w województwie łódzkim.

W roku 2002 w stosunku do roku 2001 tempo wzrostu dla Polski było nieznacznie ujemne tj. - 1,5%, podobnie jak w roku 2003, w którym wynosiło - 4,2%. W roku 2004 wszystkie województwa zanotowały dodatnie tempo zmian. Największy wzrost w 2004 roku występował w województwie lubuskim (45,7%) oraz śląskim (13,3%). Natomiast wzrost dla Polski to 7%.

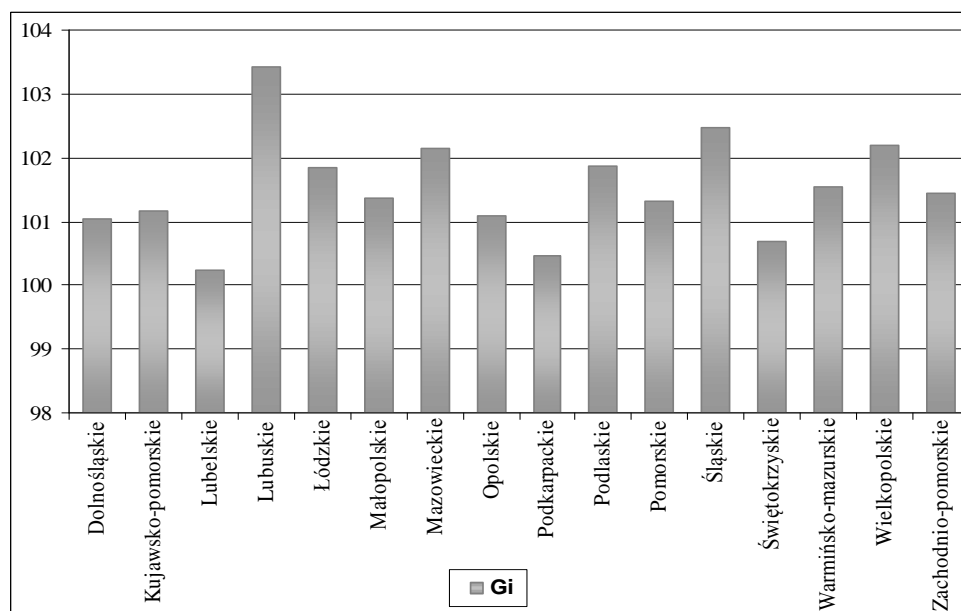
W 2005 i 2006 roku w większości województw obserwuje się ujemne tempo zmian. W 2005 roku najwyższy spadek -15,2% nastąpił w województwie lubuskim, natomiast w 2006 roku również -15,2% w województwie lubelskim.

W 2007 roku w stosunku do 2006 r., wszystkie województwa charakteryzowały się dodatnim tempem wzrostu (13,9%), najwyższe wystąpiło dla województwa śląskiego (32,9%). W 2008 roku, z wyjątkiem województwa zachodnio-pomorskiego wszystkie regiony Polski zanotowały dodatnie tempo

zmian (5,6%). Również w roku 2009 tempo wzrostu dla Polski było dodatnie i wynosiło 9,4%. Największym wzrostem w 2009 roku w stosunku do roku 2008 charakteryzowało się województwo świętokrzyskie 18,3%. W ostatnim roku badanego okresu w roku 2010 w stosunku do poprzedniego 2009 r., tempo wzrostu dla Polski było nieznacznie dodatnie i wynosiło 0,7%.

Średnie tempo zmian G_i (w %) województw dla badanego okresu ilustruje rysunek 3, natomiast tabela 3 przedstawia uporządkowanie województw ze względu na średnią wartość miernika rozwoju województw \bar{m}_i w latach 1998-2010 oraz według średniego tempa zmian G_i .

Rysunek 3. Średnie tempo zmian G_i (w %) dla województw w latach 1998-2010



Źródło: opracowanie własne

Tabela 3. Uporządkowanie województw według średniego miernika rozwoju \bar{m}_i oraz średniego tempa zmian G_i (w %)

Poz.	Województwo	\bar{m}_i	Poz.	Województwo	G_i
1	Opolskie	0,453	1	Lubuskie	103,42
2	Wielkopolskie	0,435	2	Śląskie	102,46
3	Lubelskie	0,422	3	Wielkopolskie	102,18
4	Kujawsko-pomorskie	0,413	4	Mazowieckie	102,13
5	Mazowieckie	0,411	5	Podlaskie	101,87
6	Dolnośląskie	0,381	6	Łódzkie	101,83
7	Podlaskie	0,380	7	Warmińsko-mazurskie	101,54
8	Łódzkie	0,375	8	Zachodnio-pomorskie	101,44

Poz.	Województwo	\bar{m}_i	Poz.	Województwo	G_i
9	Warmińsko-mazurskie	0,375	9	Małopolskie	101,35
10	Zachodnio-pomorskie	0,364	10	Pomorskie	101,32
11	Pomorskie	0,344	11	Kujawsko-pomorskie	101,17
12	Świętokrzyskie	0,331	12	Opolskie	101,08
13	Małopolskie	0,326	13	Dolnośląskie	101,04
14	Podkarpackie	0,285	14	Świętokrzyskie	100,69
15	Śląskie	0,284	15	Podkarpackie	100,45
16	Lubuskie	0,252	16	Lubelskie	100,24

Źródło: obliczenia własne

Wyniki zamieszczone w Tabeli 3 oraz Rysunek 3 pokazują, że województwo lubuskie, które miało najniższy średni miernik rozwoju (0,25) i zajmowało ostatnie miejsce w rankingu, charakteryzowało się jednocześnie największym średnim tempem zmian (wzrost o 3,4%). Podobnie województwo śląskie, które ze względu na średni miernik rozwoju zajmuje przedostatnią pozycję, natomiast według tempa zmian ma wysokie drugie miejsce. Województwo lubelskie natomiast odwrotnie, w rankingu zajmowało trzecią pozycję ze względu na średni poziom rozwoju, za to pod względem średniego tempa zmian ostatnie miejsce, ponieważ zanotowało niewielki wzrost (0,24%). Inną sytuację można zaobserwować w województwach podkarpackim i świętokrzyskim. Województwa te, według wyników badań charakteryzowały się niskimi średnimi poziomami rozwoju rolnictwa w latach 1998 - 2010, nie wykazując dużego tempa zmian w tym zakresie, a co za tym idzie, zajmują niskie pozycje w rankingu.

WNIOSKI

W pracy pokazano, że w rozważanym okresie kierunki i dynamika zmian poziomu rolnictwa są wyraźnie zróżnicowane pomiędzy województwami (patrz rysunek 3). Przeprowadzona analiza wielowymiarowa, przy pomocy różnych mierników syntetycznych pokazała, że regionalne zróżnicowanie poziomu rolnictwa powiększa się. Podana wyżej analiza dowodzi, że są potrzebne systematyczne badania poziomu rozwoju polskiego rolnictwa, aby można monitorować regionalne zróżnicowanie kraju pod tym względem.

LITERATURA

- Binderman A. (2004) Przestrzenne zróżnicowanie potencjału rolnictwa w Polsce w latach 1989-1998, R. N. SERIA G., T.91, Z.1, str. 51, „Wieś Jutra”, Warszawa.
- Binderman A. (2005a) Klasyfikacja polskich województw według poziomu rozwoju rolnictwa, RNR, Seria G., T.92, Z.1, str. 42, Wieś Jutra, Warszawa.
- Binderman A. (2005b) O problemie wyboru wzorca przy badaniu przestrzennego zróżnicowania potencjału rolnictwa w Polsce, MIBE – V, Warszawa, str. 46.

- Binderman A. (2006a) Wykorzystanie funkcji użyteczności do badania przestrzennego zróżnicowania rolnictwa, R. N. SERiA, Tom VIII, Zeszyt 5, Warszawa-Poznań, str.5.
- Binderman A. (2006b) Klasyfikacja obiektów oparta na dwóch wzorcach, EiOGŻ, Zeszyty Naukowe SGGW, nr 60, Warszawa, s. 25.
- Binderman A. (2007) Wielowymiarowa analiza regionalnego zróżnicowania rolnictwa w Polsce, praca doktorska, SGGW, Warszawa.
- Binderman A. (2008a) Analiza regionalnego zróżnicowania rolnictwa Polski w 2006 roku, Roczniki Naukowe SERiA, Tom X, Zeszyt 2, Warszawa-Lublin.
- Binderman A. (2008b) Zastosowanie liniowej i nieliniowej funkcji użyteczności do badania poziomu rolnictwa w Polsce, MIBE IX, wyd. SGGW, Warszawa, str. 29-38.
- Binderman A. (2009) Zależność oceny zróżnicowania rolnictwa w Polsce od wybranych mierników syntetycznych, MIBE X, wyd. SGGW, Warszawa, s. 30-41.
- Binderman A. (2010) Wpływ sposobu normalizacji zmiennych na ocenę regionalnego zróżnicowania rolnictwa, MIBE XI, 2, s. 25-38.
- Binderman A. (2011) Wielokryterialne metody analizy zróżnicowania polskiego rolnictwa w 2009 roku, MIBE XII, 2, 58-68.
- Binderman A. (2012) Regionalne zróżnicowanie rolnictwa w Polsce w latach 1998-2010, MIBE XIII/3, wyd. SGGW, Warszawa, 52-64.
- Cieślak M. (1993) Ekonomiczne zastosow. mierników syntet. ze zmiennym wzorcem, Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gosp., AE, Kraków.
- Harasim (red) (2006) Regionalne zróżnicowanie produkcji rolniczej, raporty PIB 3, Puławy.
- Hwang C. L., Yoon K. (1981) Multiple Attribute Decision Making, Methods and Applications, Springer – Verlag, New York.
- Krasowicz S. (2008) Regionalne zróżnicowanie zmian w polskim rolnictwie po integracji z Unią Europejską, Roczniki nauk Rolniczych, Seria G, T. 95, z.3/4.
- Krasowicz S. (2009) Regionalne zróżnicowanie zmian w rolnictwie polskim, Studia i raporty IUNG-PIB, z. 15, 9-37.
- Kukuła K. (2000) Metoda unitaryzacji zerowanej, PWN, Warszawa.
- Michalski T. (2004) Statystyka, WSiP, Warszawa.
- Muszyńska J. (2010) Zmiany region. Zrózn. poziomu rolnictwa w Polsce, SERiA XII.
- Urząd Komitetu Integracji Europejskiej., (2009): 5 lat w UE, www.ukie.gov.pl.

DYNAMIC OF THE REGIONAL DIFFERENTIATION OF POLISH AGRICULTURE IN THE YEARS 1998-2010

Abstract: The paper is a direct continuation of the author's research, the results of which are presented in the paper [Binderman 2012]. In the present study is using the earlier results and have been done analysis of dynamics in changes in the level of development of agriculture and its differentiation of Polish voivodeships in the years 1998-2010. The study made it possible to measure the growth of level of agriculture in Poland and to made the preparation of the classification of voivodeships due to changes in the period considered.

Keywords: synthetic measures, utility functions, agriculture development level, classification

ZASTOSOWANIE METRYKI MINKOWSKIEGO DO POMIARU ZMIAN KONCENTRACJI

Zbigniew Binderman

Państwowa Wyższa Szkoła Zawodowa w Gorzowie Wielkopolskim
e-mail: zbigniew_binderman@sggw.pl

Bolesław Borkowski

Katedra Ekonometrii i Statystyki
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
e-mail: boleslaw_borkowski@sggw.pl

Wiesław Szczesny

Katedra Informatyki
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
e-mail: wieslaw_szczesny@sggw.pl

Streszczenie: Praca jest bezpośrednią kontynuacją badań autorów i A. Prokopenyi [Binderman, Borkowski, Szczesny 2012, 2013, Binderman, Borkowski, Prokopenya, Szczesny, 2013, 2013a], dotyczących budowy nowych wskaźników koncentracji i ich stabilności. W niniejszej pracy podano nowe współczynniki koncentracji i zgodnie z wymaganymi postulatami, zbadano ich wrażliwość. Współczynniki te wykorzystują metryki Minkowskiego ρ^p ($1 \leq p \leq \infty$).

Słowa kluczowe: współczynnik koncentracji, współczynnik Giniego, współczynnik Herfindahla–Hirschmana, metryka Minkowskiego

WSTĘP

W pracy [Binderman, Borkowski, Szczesny 2012], podano sposoby wykorzystania wykresów radarowych do budowy wskaźników, służących do pomiaru koncentracji rozkładu dobra. Sposób ten wykorzystuje szeroko rozumianą ideę współczynnika Giniego, opartego o krzywą Lorenza oraz metodę przedstawioną w pracy [Binderman, Borkowski, Szczesny 2008]. Przedstawiona

technika może również posłużyć do tworzenia przez analityka nowych wskaźników koncentracji wykorzystując miary podobieństwa i niepodobieństwa obiektów, tak aby spośród stworzonych wskaźników wybrać ten, który najlepiej będzie spełniał wymagane postulaty wrażliwości [Binderman, Borkowski Szczesny 2013, Szczesny, Binderman, Prokopenya 2013, 2013a].

Pomiar koncentracji jest jednym z kluczowych zagadnień w ekonomii. Pomiar ten dokonywany jest zarówno w obszarze badań podziału dochodów jak i w obszarze badań koncentracji na poszczególnych rynkach. Literatura z tego zakresu jest bardzo bogata. Jest mnóstwo propozycji różnych wskaźników do pomiaru, należy jednak zwrócić uwagę, że wykorzystywanie tylko jednego wskaźnika, nawet powszechnie uznanego może prowadzić do przejściowego braku zauważenia zachodzących zmian. Aby ten aspekt uwypuklić, rozważmy następujący przykład pokazujący zmiany na rynku złożonym z 10-ciu podmiotów. Tabela 1 pokazuje zmiany udziałów w rynku 10-ciu podmiotów w okresie kilku kolejnych lat.

Tabela 1. Przykładowe struktury udziału w rynku dla 10 podmiotów (dane fikcyjne).

obiekt	Egalitarny	Rok 1	Rok 2	Rok 3	Rok 4	Rok 5	Rok 6	Rok 7
o01	10%	1,10%	3,0170%	1,0004%	1,98445%	5,5%	3,0%	1,8996%
o02	10%	1,11%	3,0185%	2,0000%	1,98445%	5,5%	3,8%	1,8996%
o03	10%	1,12%	3,0200%	3,0000%	1,98445%	5,5%	4,6%	1,8996%
o04	10%	1,14%	3,0250%	4,0000%	1,98445%	5,5%	5,5%	1,8996%
o05	10%	1,15%	3,0500%	6,0000%	1,98445%	5,5%	7,0%	1,8996%
o06	10%	1,16%	3,0900%	7,0000%	5,50000%	5,5%	8,1%	18,1004%
o07	10%	1,40%	3,1600%	7,9000%	7,50000%	5,5%	9,0%	18,1004%
o08	10%	29,33%	5,0950%	8,6000%	10,00000%	5,5%	10,0%	18,1004%
o09	10%	30,99%	33,0000%	11,0010%	20,00000%	5,5%	11,0%	18,1004%
o10	10%	31,50%	40,5245%	49,4986%	47,07775%	50,5%	38,1%	18,1004%
HHI	1000	2822,5	2822,5	2822,5	2822,5	2822,5	1940,7	1656,2
HHI*	0	0,2025	0,2025	0,2025	0,2025	0,2025	0,1045	0,0729
Radar	0	0,7065	0,7229	0,7151	0,7455	0,6247	0,5740	0,4651
Gini*	0	0,6940	0,6203	0,6002	0,6579	0,4500	0,4500	0,4500

Źródło: opracowanie własne

Najpopularniejszym w literaturze narzędziem do badania koncentracji rynku jest wskaźnik Herfindahla–Hirschmana (**HHI**) (ang. *Herfindahl–Hirschman Index*), określany często, jako *wskaźnik Herfindahla* lub po prostu w skrócie **HHI** [Herfindahl 1955; Hirschman, 1964]. Krótko można stwierdzić iż określa on szacunkowy poziom koncentracji w danej branży oraz poziom konkurencji na danym rynku. Jest to koncepcja ekonomiczna szeroko stosowana w prawie antymonopolowym. Nazwa współczynnika pochodzi od nazwisk dwóch ekonomistów: Orrisa Herfindahla i Alberta Otto Hirschmana, którzy go wykorzystywali. Indeks ten przez pierwszego z nich został rozpowszechniony w 1950 roku, później okazało się, że przed nim zaproponował go w 1945 roku

Hirschman. Szerzej można o wskaźniku HHI przeczytać np. w pracach [Calkins, 1983, Kwoka 1977, 1985, Matsumoto, Merlone, Szidarovszky, Djolov, Lijesen 2004, Kerber, Kretschmer, Wangenheim 2012, Ávila, Flores, López-Gallo, Márquez 2013].

Innym powszechnie stosowanym wskaźnikiem koncentracji jest wskaźnik Giniego, który jest stosowany głównie w badaniach nierówności dochodów [Barnett 2005, Hoffmann and Bradley 2007]. Oba te wskaźniki w przypadku niedużej liczby obiektów są nieunormowane, czyli w przypadku pełnego monopolu nie przyjmują wartości równej 1. Jednakże łatwo je unormować i tę unormowaną postać będziemy oznaczać gwiazdką. W przypadku normowania indeksu Giniego, wystarczy wartość wskaźnika podzielić przez wyrażenie $(n-1)/n$, gdzie n oznacza liczbę podmiotów. Natomiast przyjęta formuła dla indeksu *HHI* ma następującą postać:

$$HHI^* = \frac{HHI - \frac{1}{n}}{1 - \frac{1}{n}} \quad (1)$$

W powyższej Tabeli 1. pokazano, że przy analizach wielu rynków, ograniczając się tylko do jednego wskaźnika można nie zauważyć wielu istotnych, zachodzących zmian. Mianowicie, zmiany zachodzące w latach R1-R5 są niezauważalne przy wykorzystaniu wskaźnika *HHI**, a zmiany struktury udziałów pomiędzy latami R5, R6 i R7 są niezauważalne przy wykorzystaniu unormowanego wskaźnika Giniego. W Tabeli 1 podano również wartość innego wskaźnika koncentracji, współczynnika wykorzystującego interpretację graficzną w postaci wykresów radarowych. Jest to wskaźnik unormowany o konstrukcji podobnej do *HHI* w tym sensie, że również jemu można nadać postać relacji preferencji. Szerzej o tym wskaźniku można przeczytać w pracach [Binderman, Borkowski, Szczesny 2012, Binderman, Szczesny, Prokopenya 2013, Szczesny, Binderman, Prokopenya 2013].

Przykład zawarty w Tabeli 1 sugeruje, że w aktualnej sytuacji, gdy problemy obliczeniowe nie mają praktycznego znaczenia w badaniach powinno się wykorzystywać kilka różnych wskaźników. Dlatego w niniejszej pracy podjęto próbę konstrukcji rodziny wskaźników koncentracji w oparciu o standardowe metryki Minkowskiego [Dunford, Schwartz 1958].

WSKAŹNIK KONCENTRACJI OPARTY NA METRYCE MINKOWSKIEGO

Niech $X = \mathfrak{R}^n$ oznacza n -wymiarową przestrzeń Euklidesa, zdefiniujmy zbiory wektorów

$$\mathfrak{R}_+^n = \{\mathbf{x} = (x_1, x_2, \dots, x_n) \in X : x_i \geq 0, i = 1, 2, \dots, n\}, n \in \mathbb{N}$$

$$\Omega = \left\{ \mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathfrak{R}_+^n : \sum_{i=1}^n x_i = 1 \right\}.$$

W niniejszej pracy elementy zbioru Ω nazywać będziemy *wektorami strukturalnymi* albo krótko *strukturami*. Zbiór Ω jest częścią hiperpłaszczyzny w przestrzeni \mathfrak{R}_+^n , określonej przez równanie

$$x_1 + x_2 + \dots + x_n = 1, \quad x_i \in \mathfrak{R}_+; \quad i = 1, 2, \dots, n.$$

Indeks Herfindahla–Hirschmana (HHI): $h : \Omega \rightarrow \mathfrak{R}_+^n$, wyraża się następująco:

$$h(\mathbf{x}) = \sum_{i=1}^n (x_i)^2. \quad (2)$$

Natomiast znormalizowany indeks Herfindahla–Hirschmana (HHI*) wzorem (1).

DEFINICJA 1. Niech $\mathbf{x} = (x_1, x_2, \dots, x_n) \in X$, zbiór $A := \{x_1, x_2, \dots, x_n\} \subset X$, wektor $\mathbf{x}' := (x'_1, x'_2, \dots, x'_n)$, nazywać będziemy *porządkiem wektora \mathbf{x}* , gdzie $x'_1 \leq x'_2 \leq \dots \leq x'_n$ jest permutacją zbioru A . Wektor

$$\mathbf{x}^\wedge := (x_1^\wedge, x_2^\wedge, \dots, x_n^\wedge), \quad x_i^\wedge := \sum_{j=1}^i x'_j, \quad i = 1, 2, \dots, n,$$

nazywać będziemy *kumulacją wektora \mathbf{x}* . Jeżeli wektor \mathbf{x} jest strukturą tj. $\mathbf{x} \in \Omega$ to również wektor $\mathbf{x}' \in \Omega$, lecz wektor $\mathbf{x}^\wedge \notin \Omega$. Korzystając z powyższej definicji wektorów \mathbf{x}' i \mathbf{x}^\wedge , określimy parę operatorów \mathbf{P} i \mathbf{C} .

DEFINICJA 2. Operator $\mathbf{P} : X \rightarrow X$ nazywać będziemy **operatorem porządku** w przestrzeni X (porządkującym X), jeżeli

$$\mathbf{P}\mathbf{x} = \mathbf{x}', \quad (3)$$

dla każdego $\mathbf{x} \in X$, gdzie wektor \mathbf{x}' jest określony w Definicji 1.

DEFINICJA 3. Operator $\mathbf{C} \in L_0(X)$ nazywać będziemy **operatorem kumulacji** w przestrzeni X (kumulującym X), jeżeli

$$\mathbf{C}\mathbf{x} = \mathbf{x}^\wedge. \quad (4)$$

dla każdego $\mathbf{x} \in X$, gdzie wektor \mathbf{x}^\wedge jest określony w Definicji 1.

Zbiór wszystkich operatorów liniowych, których dziedziny są równe przestrzeni X (podzbiorem liniowymi X) i których wartości należą również do przestrzeni X oznaczymy przez $L_0(X)$ ($L(X)$) [Przeworska–Rolewicz 1977, Przeworska–Rolewicz, Rolewicz 1968].

Z definicji operatora \mathbf{P} wynika bezpośrednio następujące twierdzenie.

TWIERDZENIE 1. Operator $P \in L_0(\Omega)$.

Operator $C \notin L_0(\Omega)$, świadczy o tym następujący przykład.

PRZYKŁAD 1. Niech $\mathbf{x} = (0, 1; 0, 4; 0, 5)$, $\mathbf{y} = (0, 5; 0, 4; 0, 1)$. Wówczas

$$\mathbf{x} + \mathbf{y} = (0, 6; 0, 8; 0, 6), \quad C\mathbf{x} = \mathbf{x}, \quad C\mathbf{y} = \mathbf{x}.$$

Stąd,

$$C\mathbf{x} + C\mathbf{y} = 2\mathbf{x} = (0, 2; 0, 8; 1) \neq C(\mathbf{x} + \mathbf{y}).$$

DEFINICJA 4. Niech $\mathbf{x} = (x_1, x_2, \dots, x_n) \in X$, operator

$$S_\varepsilon^{(j,i)} : \mathfrak{R}_+ \times X \rightarrow X, \quad j \in \{1, 2, \dots, n-1\}, \quad i : 1 \leq i \leq n-j, \quad \varepsilon \geq 0,$$

określony za pomocą wzoru

$$S_\varepsilon^{(j,i)}(\mathbf{x}) = (x_1, x_2, \dots, x_{j-1}, x_j - \varepsilon, x_{j+1}, \dots, x_{j+i-1}, x_{j+i} + \varepsilon, x_{j+i+1}, \dots, x_n), \quad (5)$$

nazywać będziemy *operatorem przesunięcia (shift operator)*.

Niech wektory $\mathbf{x} = (x_1, x_2, \dots, x_n)$, $\mathbf{y} = (y_1, y_2, \dots, y_n) \in X$. W przestrzeni Euklidesa, jako miary zróżnicowania wektorów, przeważnie wykorzystywane są miary odległości typu Minkowskiego:

$$d_p(\mathbf{x}, \mathbf{y}) = \left[\sum_{j=1}^n |x_j - y_j|^p \right]^{\frac{1}{p}}, \quad \infty \geq p \geq 1, \quad (6)$$

Do tej grupy należą poniższe metryki.

- **Metryka miejska (Hamminga, Manhattan).** Ta odległość jest sumą różnic mierzonych wzdłuż wymiarów. W większości przypadków ta miara odległości daje podobne wyniki, jak zwykła odległość euklidesowa. Jednak w przypadku tej miary, wpływ pojedynczych dużych różnic (przypadków odstających) jest stłumiony, ponieważ nie podnosi się ich do kwadratu. Odległość miejską oblicza się następująco:

$$d_1(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^n |x_j - y_j|. \quad (7)$$

- **Metryka Euklidesa.** Należy do jednej z najczęściej wybieranych typów odległości. Jest to odległość geometryczna w przestrzeni wielowymiarowej. Oblicza się ją następująco:

$$d_2(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{j=1}^n (x_j - y_j)^2}, \quad (8)$$

Odległości euklidesowe (oraz kwadraty odległości euklidesowych) są wyliczane w oparciu o surowe dane, a nie na podstawie danych standaryzowanych. Metoda ta ma pewne zalety (np. na odległość pomiędzy dwoma dowolnymi obiektami nie ma wpływu dodanie do analizy nowych obiektów, które mogą być obiektami odstającymi). Jednakże na odległości duży wpływ mają różnice jednostek między wymiarami, na podstawie których są wyliczane odległości. Może to spowodować otrzymanie zupełnie innych wyników.

- **Odległość Czebyszewa.** Taka miara odległości jest stosowna w tych przypadkach, w których chcemy zdefiniować dwa obiekty jako "inne" wtedy, gdy różnią się one w jednym dowolnym wymiarze.

$$d_{\infty}(\mathbf{x}, \mathbf{y}) = \max_{1 \leq j \leq n} |\mathbf{x}_j - \mathbf{y}_j|, \quad (9)$$

DEFINICJA 5. Niech wektory $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \Omega$, $\frac{\mathbf{1}}{\mathbf{n}} = \left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right) \in \Omega$,

$\mathbf{s} = (0, \dots, 0, 1) \in \Omega$, d_p będzie metryką Minkowskiego określoną przez wzór (5), operatory \mathbf{P} , \mathbf{C} są określone przez definicje 2, 3, odpowiednio. Funkcje $m, M: \Omega \rightarrow \mathfrak{R}_+$, określoną za pomocą wzoru:

$$m_p(\mathbf{x}) = \alpha_p d_p \left(\mathbf{C} \left(\frac{\mathbf{1}}{\mathbf{n}} \right), \mathbf{C} \mathbf{P} \mathbf{x} \right), \quad 1 \leq p \leq \infty, \quad (10)$$

$$M_p(\mathbf{x}) = d_p \left(\frac{\mathbf{1}}{\mathbf{n}}, \mathbf{P} \mathbf{x} \right) / d_p \left(\frac{\mathbf{1}}{\mathbf{n}}, \mathbf{s} \right), \quad 1 \leq p \leq \infty, \quad (10a)$$

gdzie

$$\alpha_p = n \left[\sum_{j=1}^{n-1} j^p \right]^{-\frac{1}{p}}, \quad 1 \leq p < \infty, \quad \alpha_{\infty} = \frac{n}{n-1}, \quad (11)$$

nazywać będziemy *miarami koncentracji* wektora \mathbf{x} .

UWAGA 1. Można pokazać, że

$$\alpha_1 = \frac{2}{n-1}, \alpha_2 = \sqrt{\frac{6n}{(n-1)(2n-1)}}, \alpha_3 = \sqrt[3]{\frac{4n}{(n-1)^2}}, \alpha_4 = \sqrt[4]{\frac{30n^3}{(n-1)(2n-1)(3n^2-3n-1)}}.$$

W pracy [Binderman, Borkowski, Szczesny 2012] udowodniono następujący lemat.

LEMAT 1. Jeżeli wektor $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \Omega$ to $\mathbf{C} \left(\frac{\mathbf{1}}{\mathbf{n}} \right) \geq \mathbf{C} \mathbf{P} \mathbf{x}$.

Łatwo sprawdzić, że zdefiniowany wyżej miernik ma własności.

WŁASNOŚĆ 1. Miernik koncentracji m wektorów strukturalnych (należących do Ω) ma następujące własności:

- (i) $0 \leq m(\mathbf{x}) \leq 1$, $\mathbf{x} \in \Omega$, (ii) $m\left(\frac{\mathbf{1}}{\mathbf{n}}\right) = 0$, (iii) $m(\mathbf{x}) = 1$ dla wektorów osobliwych,

gdzie przez wektor *osobliwy* rozumiemy wektor należący do Ω , którego jedna ze współrzędnych jest równa 1.

LEMAT 2. Niech $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \Omega$, $\mathbf{P}\mathbf{x} = \mathbf{x}' = (x'_1, x'_2, \dots, x'_n)$, $\varepsilon > 0$.

Jeżeli $\mathbf{x}_\varepsilon = \mathbf{S}_\varepsilon^{(j,1)}(\mathbf{P}\mathbf{x})$ to

$$m(\mathbf{x}_\varepsilon) \geq m(\mathbf{x}). \quad (12)$$

DOWÓD. Z definicji mamy

$\mathbf{S}_\varepsilon^{(j,1)}(\mathbf{P}\mathbf{x}) = \mathbf{S}_\varepsilon^{(j,1)}(\mathbf{x}') = (x'_1, x'_2, \dots, x'_{j-1}, x'_j - \varepsilon, x'_{j+1} + \varepsilon, x'_{j+2}, \dots, x'_n)$ oraz

$$\mathbf{C}\mathbf{x}_\varepsilon = \mathbf{x}^\wedge = (x_1^\wedge, x_2^\wedge, \dots, x_{j-1}^\wedge, x_j^\wedge - \varepsilon, x_{j+1}^\wedge, x_{j+2}^\wedge, \dots, 1)$$

$$\mathbf{x}^\wedge = \mathbf{C}\mathbf{x}' = (x_1^\wedge, x_2^\wedge, \dots, x_{n-1}^\wedge, 1); \frac{\mathbf{1}^\wedge}{\mathbf{n}} = \left(\frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1\right), \quad (13)$$

$$\mathbf{x}_\varepsilon^\wedge = \mathbf{C}\mathbf{x}'_\varepsilon = (x_1^\wedge, x_2^\wedge, \dots, x_{j-1}^\wedge, x_j^\wedge - \varepsilon, x_{j+1}^\wedge, \dots, x_{n-1}^\wedge, 1). \quad (14)$$

Stąd,

$$\begin{aligned} m_p(\mathbf{x}) &= \alpha_p d_p \left(\mathbf{C} \left(\frac{\mathbf{1}}{\mathbf{n}} \right), \mathbf{C}\mathbf{P}\mathbf{x} \right) = \alpha_p d_p \left(\frac{\mathbf{1}^\wedge}{\mathbf{n}}, \mathbf{x}^\wedge \right) = \alpha_p \left[\sum_{i=1}^n \left(\frac{i}{n} - x_i^\wedge \right)^p \right]^{1/p} = \\ &= \alpha_p \left[\sum_{i=1}^{j-1} \left(\frac{i}{n} - x_i^\wedge \right)^p + \left(\frac{j}{n} - x_j^\wedge \right)^p + \sum_{i=j+1}^n \left(\frac{i}{n} - x_i^\wedge \right)^p \right]^{1/p}, \\ m_p(\mathbf{x}_\varepsilon) &= \alpha_p d_p \left(\frac{\mathbf{1}^\wedge}{\mathbf{n}}, \mathbf{x}_\varepsilon^\wedge \right) = \alpha_p \left[\sum_{i=1}^{j-1} \left(\frac{i}{n} - x_i^\wedge \right)^p + \left(\frac{j}{n} + \varepsilon - x_j^\wedge \right)^p + \sum_{i=j+1}^n \left(\frac{i}{n} - x_i^\wedge \right)^p \right]^{1/p} \end{aligned}$$

Na podstawie dwóch powyższych równości wnioskujemy, że

$$m(\mathbf{x}_\varepsilon) - m(\mathbf{x}) \geq 0.$$

UWAGA 2. Można pokazać że jeżeli $\mathbf{y}, \mathbf{z} \in \Omega$ to

$$\mathbf{C}\mathbf{P}\mathbf{z} \geq \mathbf{C}\mathbf{P}\mathbf{y} \Rightarrow m(\mathbf{y}) \geq m(\mathbf{z}). \quad (15)$$

UWAGA 3. Łatwo zauważyć, że dla wektorów $\mathbf{x}^\wedge, \mathbf{x}_\varepsilon^\wedge$ słuszna jest nierówność

$$\mathbf{x}^\wedge \geq \mathbf{x}_\varepsilon^\wedge,$$

gdzie wektory $\mathbf{x}^\wedge, \mathbf{x}_\varepsilon^\wedge$ są określone za pomocą operacji (4), (5), odpowiednio.

Stąd na podstawie Uwagi 2. można otrzymać tezę Lematu 2.

Korzystając z Lematu 2 lub Uwag 2,3 w podobny sposób, jak w pracy [Binderman, Szczesny, Prokopenya 2013a –dowód Wniosku 2] można udowodnić twierdzenie dotyczące wartości operatora $\mathbf{S}_\varepsilon^{(j,i)}(\mathbf{P}\mathbf{x})$.

TWIERDZENIE 2. Jeżeli wektor $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \Omega$, wektor $\mathbf{x}_\varepsilon = \mathbf{S}_\varepsilon^{(j,i)}(\mathbf{P}\mathbf{x}) \in \Omega$, liczba $\varepsilon > 0$, to

$$m(\mathbf{x}_\varepsilon) \geq m(\mathbf{x}).$$

TWIERDZENIE 3. Dla dowolnego $p: 1 \leq p \leq \infty$, $\mathbf{x} \in \Omega$ współczynnik koncentracji indukowany przez metryki typu Minkowskiego

$$m_p(\mathbf{x}) = \alpha_p d_p \left(\mathbf{C} \left(\frac{\mathbf{1}}{\mathbf{n}} \right), \mathbf{C}\mathbf{P}\mathbf{x} \right),$$

gdzie d_p jest metryką określoną przez wzór (4), α_p jest określone przez wzór (11),

wektor $\frac{\mathbf{1}}{\mathbf{n}} = \left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n} \right) \in \Omega$, operatory \mathbf{P}, \mathbf{C} są określone przez definicje 2, 3,

odpowiednio, spełnia wszystkie (cztery) podstawowe postulaty stawiane miarom koncentracji w badaniu nierówności [Binderman, Borkowski, Prokopenya, Szczesny 2013, 2013a].

PRZYKŁAD 2. Dla wektora $\mathbf{x}_\varepsilon = \frac{\mathbf{1}}{\mathbf{n}_\varepsilon} = \left(\frac{1}{n} - \varepsilon, \frac{1}{n}, \dots, \frac{1}{n} + \varepsilon \right), 0 < \varepsilon \leq \frac{1}{n}$

otrzymujemy:

$$m_p(\mathbf{x}_\varepsilon) = \alpha_p \left[\sum_{i=1}^{n-1} \varepsilon^p \right]^{1/p} = \alpha_p (n-1)^{1/p} \varepsilon.$$

UWAGA 4. W sposób podobny do powyższego można udowodnić analogony Twierdzeń 2,3 dla współczynnika koncentracji M_2 , określonego w Definicji 5.

UWAGA 5. Podany we wzorze (2) znormalizowany indeks Herfindahla-Hirschmana można określić za pomocą metryki Euklidesa ($p=2$)

$$h^*(\mathbf{x}) = \frac{d_2^2\left(\frac{\mathbf{1}}{\mathbf{n}}, \mathbf{x}\right)}{d_2^2\left(\frac{\mathbf{1}}{\mathbf{n}}, \mathbf{s}\right)} = M_2^2(\mathbf{x}), \quad \mathbf{x}, \mathbf{s} \in \Omega,$$

gdzie \mathbf{s} jest dowolnym wektorem osobliwym, metryka d_2 – jest określona za pomocą wzoru (8), miernik koncentracji M jest określony za pomocą wzoru (10a).

Istotnie, niech wektory

$\mathbf{x}, \mathbf{s} \in \Omega$: $\mathbf{x} = (x_1, x_2, \dots, x_n)$, $\mathbf{s} = (0, 0, \dots, 0, 1, 0, 0, \dots, 0)$ będą dowolnie ustalone, mamy wówczas

$$\begin{aligned} \frac{d_2^2\left(\frac{\mathbf{1}}{\mathbf{n}}, \mathbf{x}\right)}{d_2^2\left(\frac{\mathbf{1}}{\mathbf{n}}, \mathbf{s}\right)} &= \frac{\sum_{i=1}^n \left(\frac{1}{n} - x_i\right)^2}{\sum_{i=1}^{n-1} \left(\frac{1}{n^2}\right) + \left(1 - \frac{1}{n}\right)^2} = \frac{\sum_{i=1}^n \left(\frac{1}{n}\right)^2 + \sum_{i=1}^n (x_i)^2 - 2\frac{1}{n} \sum_{i=1}^n x_i}{\sum_{i=1}^n \left(\frac{1}{n^2}\right) + 1 - \frac{2}{n}} = \\ &= \frac{\sum_{i=1}^n x_i^2 - \frac{1}{n}}{1 - \frac{1}{n}} = h^*(\mathbf{x}). \end{aligned}$$

W pracach wielu analityków podawane są wartości koncentracji określonego dobra, będącego elementem rozważanych obiektów, w porównywanych populacjach. Najczęściej do takich porównań używa się współczynnika koncentracji Giniego. Niech $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \Omega$,

$$\mathbf{x}' = \mathbf{P}\mathbf{x} = (x'_1, x'_2, \dots, x'_n), \quad \mathbf{x}^\wedge = \mathbf{C}\mathbf{P}\mathbf{x} = (x^\wedge_1, x^\wedge_2, \dots, x^\wedge_n),$$

gdzie operatory \mathbf{P} i \mathbf{C} są określone za pomocą wzorów (3), (4), odpowiednio. Unormowany współczynnik Giniego można również obliczyć za pomocą wzoru:

$$G^*(\mathbf{x}) = 1 - \frac{2}{(n-1)} \sum_{i=1}^{n-1} x_i^\wedge. \quad (16)$$

UWAGA 6. Podany we wzorze (16) współczynnik Giniego można wyrazić za pomocą metryki miejskiej ($p=1$)

$$G^*(\mathbf{x}) = \frac{d_1\left(\frac{\mathbf{1}^\wedge}{\mathbf{n}}, \mathbf{x}^\wedge\right)}{d_1\left(\frac{\mathbf{1}^\wedge}{\mathbf{n}}, \mathbf{s}^\wedge\right)}, \quad \mathbf{x}, \mathbf{s} \in \Omega,$$

gdzie \mathbf{s} jest dowolnym wektorem osobliwym, metryka d_1 – jest określona za pomocą wzoru (7).

Istotnie, niech wektory

$\mathbf{x}, \mathbf{s} \in \Omega$: $\mathbf{x} = (x_1, x_2, \dots, x_n)$, $\mathbf{s} = (0, 0, \dots, 0, 1, 0, 0, \dots, 0)$ będą dowolnie ustalone, korzystając z nierówności [Binderman, Borkowski, Szczesny 2012]

$$x_k^{\wedge} \leq \frac{k}{n}, \quad k = 1, 2, \dots, n,$$

mamy wówczas

$$\frac{d_1\left(\frac{\mathbf{1}^{\wedge}}{\mathbf{n}}, \mathbf{x}^{\wedge}\right)}{d_1\left(\frac{\mathbf{1}^{\wedge}}{\mathbf{n}}, \mathbf{s}^{\wedge}\right)} = \frac{\sum_{i=1}^{n-1} \left(\frac{i}{n} - x_i^{\wedge}\right)}{\sum_{i=1}^{n-1} \frac{i}{n}} = \frac{\frac{(n-1)n}{2n} - \sum_{i=1}^{n-1} x_i^{\wedge}}{\frac{(n-1)n}{2n}} = 1 - \frac{2}{(n-1)} \sum_{i=1}^{n-1} x_i^{\wedge} = G^*(\mathbf{x}).$$

PODSUMOWANIE

W poprzednim podrozdziale pokazano, że dwa popularne wskaźniki w wersji unormowanej (Gini* oraz HHI*) można uzyskać przy wykorzystaniu metryki Minkowskiego. Reagują one w różny sposób na zmianę wielkości odstępstwa od rozkładu egalitarnego. Dlatego wykorzystywane są one na ogół do różnych zastosowań, a mianowicie Gini* jest wykorzystywany do badania nierówności w dostępie do określonego dobra, a HHI* do oceny koncentracji na rynku. Wykorzystując zaprezentowane w tej pracy idee konstrukcji wskaźników, przy wykorzystaniu metryki Minkowskiego możemy dobrać wskaźnik w ten sposób, aby mocno reagował na konkretne zmiany koncentracji (odstępstwa od rozkładu egalitarnego). Zauważmy ponadto, że jeśli W jest unormowanym wskaźnikiem koncentracji o wartościach z przedziału $[0;1]$, to W^a , $a > 0$ też może służyć do pomiaru koncentracji. Zatem, stosując tylko ten dodatkowy mechanizm mamy duże możliwości w zakresie konstrukcji nowych mierników koncentracji.

Popularnym sposobem przedstawienia koncentracji jest krzywa Lorenza. Rozważmy pewną rodzinę krzywych, która daje się przedstawić za pomocą poniższego wzoru (17). Rozpatrzmy ją, jako model odniesienia dla zilustrowania zmian wartości wskaźników w zależności od zmian koncentracji:

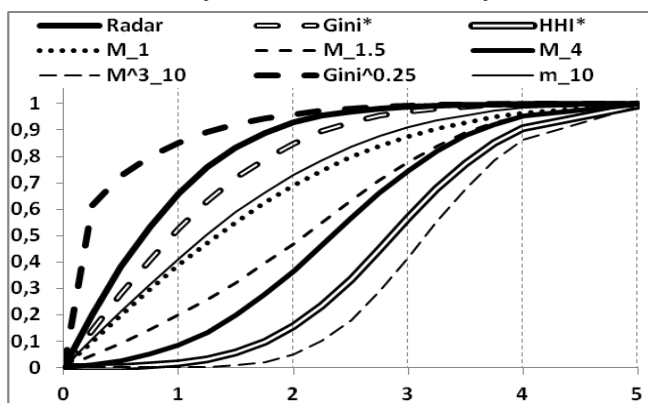
$$L(t) = F_{\mu}(F_0^{-1}(t)), \quad (17)$$

gdzie F_{μ} oznacza pewną dystrybuantę rozkładu normalnego ze średnią μ i odchyleniem standardowym 1.

Na podanym niżej Rys. 1 zilustrowano wykresy wartości wybranych wskaźników w zależności od stopnia koncentracji w wybranej rodzinie rozkładów w przedziale $[0;1]$, ilustrują one koncentrację, wyrażoną przez rodzinę krzywych,

określoną wzorem (17). Zwróćmy uwagę na zupełnie inne reakcje na zmianę koncentracji w zakresie małych i dużych poziomów koncentracji dla wskaźników pomiaru koncentracji GINI* oraz HHI*. Z omawianego rysunku widać także, że wykorzystując te idee budowy, można uzyskać wskaźniki o różnych wrażliwościach na zmiany poziomu koncentracji. Oczywiście, badacz przy wyborze wskaźnika powinien rozpatrzyć także inne parametryczne rodziny, ilustrujące zmiany koncentracji i prześledzić wrażliwość wybieranego narzędzia do pomiaru. Powinien to być model lub modele, które najlepiej pasują do zjawiska, które mogą śledzić konstruowany wskaźnik. Wtedy z całej palety możliwości można wybrać taki wskaźnik lub lepiej dwa różne wskaźniki, które mocno reagują na poziomie koncentracji, postulowanym przez użytkownika.

Rysunek 1. Wykresy wartości wybranych wskaźników koncentracji w zależności od wartości parametru μ indeksującego zmianę koncentracji według modelu (17), (na rysunku przyjęto następujące oznaczenia techniczne: M_p jako M_p , m_{10} jako m_p oraz W^a_p zamiast $(W_p)^a$, gdzie W jest nazwą wskaźnika.



Źródło: opracowanie własne

BIBLIOGRAFIA

- Ávila F., Flores E., López-Gallo F., Márquez J. (2013) Concentration indicators: Assessing the gap between aggregate and detailed data. IFC Bulletin No 36, Statistical issues and activities in a changing environment, 542-559.
- Barnett R. A. et al. (2005) College Mathematics for Business, Economics, Life Sciences, and Social Sciences, 10th ed., Prentice-Hall, Upper Saddle River.
- Binderman A. (2006) Klasyfikacja danych na podstawie dwóch wzorców, *Ekonomika i Organizacja Gospodarki Żywnościowej*, SGGW, Warszawa, z. 60, s.25-34.
- Binderman A. (2011) Wielokryterialne metody analizy zróżnicowania polskiego rolnictwa w 2009 roku, *Metody ilościowe w badaniach ekonomicznych*, XII/2, s. 58-68.
- Binderman Z., Borkowski B., Szczesny W. (2012) Radar coefficient of concentration, *Quantitative methods in economics Vol. XIII*, nr 2, 7-21.

- Binderman Z., Borkowski B., Szczesny W. (2008) O pewnej metodzie porządkowania obiektów na przykładzie regionalnego zróżnicowania rolnictwa, *Metody ilościowe w badaniach ekonomicznych*, IX, 39-48, wyd. SGGW.
- Binderman Z., Szczesny W. (2009) Arrange methods of tradesmen of software with a help of graphic representations Computer algebra systems in teaching and research, Siedlce, wyd. WSFiZ, 117-131.
- Binderman Z., Borkowski B., Prokopenya A., Szczesny W. (2013) Mechanisms of Construction of the Radar Coefficients of Concentration, *Computer Algebra Systems in Teaching and Research*, Siedlce, ed. Collegium Mazovia, 7-15.
- Binderman Z., Borkowski B., Prokopenya A., Szczesny W. (2013a) Radar Coefficients of Concentration, Verifications of Properties, *Computer Algebra Systems in Teaching and Research*, Siedlce, ed. Collegium Mazovia, 16-28.
- Binderman Z. (2011) Matematyczne aspekty metod radarowych, *Metody ilościowe w badaniach ekonomicznych*, XII, nr 2, 69-79.
- Calkins, S. (1983) The New Merger Guidelines and the Herfindahl-Hirschman Index, *California Law Review* 71 (2), 402-429.
- Djolov G.: The Herfindahl-Hirschman: Index as a decision tool to business concentration in the banking sector: a general perspective, https://umsbe.wufoo.com/cabinet/276/298/53/p171_update3.docx.
- Dunford N., Schwartz J. T. (1958) *Linear operators*, Yale University, Interscience Publishers, New York, London.
- Gelfand I., M. (1961) *Lectures on linear algebra*, Interscience, New York.
- Herfindahl, O.C. (1955) Comment on Rosenbluth's measures of concentration, *Business concentration and price policy*, ed. G. Stigler, Princeton University Press, USA.
- Hirschman, A.O. (1964) The paternity of an index", *American Economic Review*, vol. 54, no. 5, pp. 761-762.
- Hoffmann . L. D. and Bradley G. L. (2007) *Calculus for Business, Economics, and the Social and Life Sciences*, 9th ed., McGraw Hill, New York.
- Kerber W., Kretschmer J-P, Wangenheim G. (2009) Market share thresholds and Herfindahl-Hirschman-Index (HHI) as screening instruments in competition law: A theoretical analysis. discussion paper.
- Hwang C., L., Yoon K. (1981) *Multiple Attribute Decision Making: Methods and Applications*, Springer Verlag, New York.
- Kwoka, J. E. (1977) Large firm dominance and price-cost margins in manufacturing industries." *Southern Econ J.* vol. 44, pp. 183-189.
- Kwoka, J. E. (1985) The Herfindahl index in theory and practice, *Antitrust Bulletin* 30, 915-947.
- Lijesen M. G. (2004) Adjusting the Herfindahl index for close substitutes: an application to pricing in civil aviation *Transportation Research Part E* 40 123-134.
- Matsumoto A., Merlone U, Szidarovszky F. (2012) Some notes on applying the Herfindahl - Hirschman index *Applied Economics Letters*, Taylor and Francis Journals, vol. 19(2), 181-184.
- Przeworska-Rolewicz D. (1977) *Przestrzenie liniowe i operatory liniowe*. WNT Warszawa.
- Przeworska-Rolewicz D., Rolewicz S. (1968) *Equations in linear spaces*, *Monografie Matematyczne* 47, PWN Warszawa.

MODELE PŁAC TYPU MINCERA NA PRZYKŁADZIE PRACOWNIKÓW NAUKOWO-DYDAKTYCZNYCH POLSKICH WYŻSZYCH UCZELNI PUBLICZNYCH

Mariusz Franciszek Kaszubowski

Katedra Nauk Ekonomicznych, Politechnika Gdańska

e-mail: mkaszubo@zie.pg.gda.pl

Streszczenie: Artykuł przedstawia wyniki badań, których celem było wyznaczenie oraz zestawienie modeli płac typu Mincera dla grupy pracowników naukowo-dydaktycznych polskich wyższych uczelni publicznych. Dane wykorzystane do analizy pochodziły z ankiety ogólnopolskiej przeprowadzonej na wiosnę 2012 roku. Porównywane modele dotyczyły wynagrodzenia faktycznie otrzymywanego na a także poza uczelnią macierzystą oraz wynagrodzenia progowego dla hipotetycznego kontraktu wprowadzającego warunek jednoetatowości.

Słowa kluczowe: płaca progowa, model płac typu Mincera, wynagrodzenia nauczycieli akademickich

WSTĘP

Celem badania było wyznaczenie oraz porównanie modeli płac typu Mincera dla grupy pracowników naukowo-dydaktycznych polskich wyższych uczelni publicznych. Konfrontacji poddano modele wynagrodzenia faktycznie otrzymywanego z modelami wynagrodzenia progowego (reservation wage) dla hipotetycznego kontraktu wprowadzającego warunek jednoetatowości. Do analizy wykorzystano dane otrzymane w wyniku przeprowadzonej ankiety ogólnopolskiej.

W literaturze pojawia się wiele analiz dotyczących modeli płac typu Mincera dla różnych grup zawodowych, w tym również dla nauczycieli akademickich [Sadeghi i in. 2012] jednak nie sposób znaleźć tam tychże modeli dla płacy progowej wyznaczonej przy warunku ograniczenia wieloetatowości. Dodatkowo poniższa analiza daje obraz oczekiwań środowiska akademickiego w kontekście wchodzącej reformy Szkolnictwa Wyższego w Polsce.

PRZEGLĄD LITERATURY

Pierwszy model, w którym pojawia się zależność pomiędzy wynagrodzeniem a sumą wiedzy zdobytej w szkole i doświadczenia z pracy pojawił się w modelu kapitału ludzkiego zaproponowanego przez Mincera [Mincer 1958]. Równanie, uznane później za tradycyjne, w którym zależność logarytmu płac wyrażała funkcja kwadratowa zdobytego doświadczenia pojawia się nieco później [Mincer 1974]. W międzyczasie również w pracy [Schultz 1961] można było znaleźć argumentację, według której główną przyczyną wzrostu wynagrodzenia jest inwestycja w kapitał ludzki. Od tego momentu pojawiać zaczęły się kolejne modyfikacje tego modelu. O różnicach w wynagrodzeniu wynikające ze zmiany produktywności w zależności od stażu pracy pisał [Becker 1962]. On też w swojej pracy tłumaczył zmniejszający się wzrost zarobków wraz z wiekiem czy też zwiększoną liczbę szkoleń w pracy wśród młodych pracowników. Kolejną wersją równania Mincera wprowadził [Ben-Porath 1967]. Jego model oparty był na funkcji produkcji kapitału ludzkiego i zakładał maksymalizację wartości bieżącej sumy swoich zarobków. Wykorzystując teorię optymalnego sterowania [Sheshinski 1968] wyznaczał ekstrema zaproponowanych wcześniej funkcji. Wykorzystując tą samą metodę podobne badania prowadził [Haley 1973]. Kilka lat później w swojej propozycji modelu płac [Ryder i in. 1976] dodatkowo uwzględnili wpływ wolnego czasu. Jeszcze w tym samym roku [Haley 1976] zastosował metodę wyznaczania współczynników dla regresji nieliniowej. [Leibowitz 1976] pokazał natomiast, że intensywność samokształcenia ma istotny wpływ na kształt krzywej Mincera. [Driffill 1980] dodatkowo uwzględnił w swoich rozważaniach wpływ wieku przejścia na emeryturę. W równaniu zaproponowanym przez [Behrman i Birdsall 1983] pojawia się natomiast podejście polegające na rozpatrywaniu stopy zwrotu z inwestycji w edukację jako funkcji jakości początkowego stadium nauki. Postać omawianej zależności weryfikowali również [Heckman i Polachek 1974] oraz [Frazis i Loewenstein 2005].

Równanie Mincera całkiem dobrze spisywało się w wieloletniej historii empirycznych badań wysokości płac w zależności od poziomu edukacji i stażu pracy [Lemieux 2003, Heckman 2003, Heckman 2006]. Kwadratowa postać zależności logarytmu płacy od doświadczenia często jednak niedoszacowuje wartości płac w początkowym okresie i przeszacowuje je w późniejszym okresie w szczególności dla pracowników wysoko wykwalifikowanych. Wielomian trzeciego stopnia lepiej obrazuje wspomnianą zależność jeżeli jednostka badanej społeczności ma możliwość wyboru optymalnego poziomu kontynuacji nauki przez cały swój okres pracy. Pisali o tym [Murphy i Welch 1990] zaś ostatnio uzasadnił przedstawiając dowód i interpretacje [Hamlen 2012]. Uznali oni, że pewne grupy zawodowe przykładowo prawnicy, lekarze, księgowi borykają się z podjęciem decyzji czy warto jeszcze inwestować w dalszy swój rozwój (który zapoczątkuje w przyszłości). Rozwój w rozumieniu ponad minimum niezbędne by

pozostać na stanowisku. W podobnej sytuacji są polscy nauczyciele akademicki. Taki stan rzeczy powoduje, że marginalna użyteczność dochodu pomniejszonego o nakłady na kontynuację nauki jest malejącą funkcją czasu. Innymi słowy, funkcja określająca udział obecnego wynagrodzenia poświęconego na kontynuowanie edukacji jest funkcją malejącą i wypukłą ku górze a nie jak zakładano dotychczas w tradycyjnym modelu płac typu Mincera - liniową zależnością. Na koniec warto wspomnieć, że w wielu opracowaniach model płac typu Mincera wzbogacany był również o nieliniową zależność logarytmu płac od poziomu edukacji [Mincer 1997, Card i Krueger 1992, Heckman 1996, Deschênes 2001].

MATERIAŁ I METODY

Materiał statystyczny, który posłużył do wyznaczenia modeli płac pochodził z ankiety przeprowadzonej w dniach 28 lutego do 30 kwietnia 2012 roku. Populacja objęta badaniem to wszyscy pracownicy naukowo-dydaktyczni polskich wyższych uczelni publicznych. Grupa ta liczyła 51334 osób (stan na okres 2010/2011) [GUS, 2011]. Badanie miało charakter całościowy a decyzja o uczestnictwie w badaniu leżała w gestii dziekanów poszczególnych wydziałów. Respondentów pytano drogą elektroniczną o obecnie otrzymywane wynagrodzenie na uczelni macierzystej jak i poza nią a także o najniższą akceptowalną płacę dla hipotetycznego kontraktu wprowadzającego warunek jednoetatowości. Otrzymałą próbę oczyszczono z obserwacji błędnie wypełnionych oraz o wartościach odstających. Pozostałym nadano odpowiednie wagi względem płci, rodzaju uczelni i zajmowanego stanowiska.

Ostatnim zabiegiem było przeliczenie wynagrodzenia progowego z netto na brutto oraz wystandaryzowanie wynagrodzenia pod względem pensum ustalając je na wysokość 180 godzin.

Tabela 1. Charakterystyka populacji generalnej i próby

Grupa	Liczba osób	
	Populacja	Próba
Pracownicy naukowo-dydaktyczni polskich wyższych uczelni publicznych	51334	1618
Pracownicy zatrudnieni na stanowisku asystenta	6937	253
Pracownicy zatrudnieni na stanowisku adiunkta	28520	978
Pracownicy zatrudnieni na stanowisku adiunkta z habilitacją	2331	71
Pracownicy zatrudnieni na stanowisku profesora nadzwyczajnego	9287	222
Pracownicy zatrudnieni na stanowisku profesora zwyczajnego	4272	94
Kobiety	20477	670
Mężczyźni	30857	948

Źródło: opracowanie własne na podstawie ankiety oraz GUS

Do oszacowania modeli płac zastosowano najpopularniejszą Metodę Najmniejszych Kwadratów. Alternatywą mogła być proponowana po raz pierwszy przez [Bierens 1997] Metoda Momentów Ważonych, lecz ponieważ daje ona podobne wyniki co wcześniej wspomniana mniej skomplikowana MNK podobnie jak większość badaczy [m.in. Kot 1999] zastosowano metodę prostszą numerycznie.

Weryfikacja założenia homoscedastyczności elementu losowego dla modeli regresji wielorakiej sprawdzono testem White'a, gdyż jest on stosowany również dla modeli nieliniowych. Założenie normalności zweryfikowano poddając ocenie graficznej histogramy reszt. Zdecydowano się na takie podejście z dwóch powodów. W sytuacji dużych prób większość znanych testów normalności nakazują odrzucenie hipotezy zerowej mimo zadawalającego dopasowania oraz testy służące do weryfikacji istotności modelu i poszczególnych jego współczynników (test t i F) są odporne na niewielkie odchylenia od normalności [Greene 2002].

W prezentowanej analizie pojawiają się osobno modele dla dwóch zmiennych zależnych. Pierwsza to miesięczne, wystandaryzowane wynagrodzenie progowe, brutto dla hipotetycznego kontraktu przy założeniu pensum 180 godzin zwane później wynagrodzeniem progowym. Druga zmienna to średni, miesięczny, łączny, faktyczny dochód brutto (z tytułu pracy na i poza uczelnią macierzystą) nazwany dochodem. Statystyczna istotność przedstawionych modeli jak i ich poszczególnych zmiennych objaśniających zostały zweryfikowane dla poziomu istotności $\alpha = 0,05$.

W modelu płac typu Mincera pojawia się zależność logarytmu płac od wykształcenia jak i zdobytego doświadczenia. Ponieważ w przypadku grupy pracowników naukowo-dydaktycznych wszyscy ukończyli studia wyższe za zmienną określającą poziom wykształcenia uznano zajmowane stanowisko. Za zmienną określającą zdobyte doświadczenie wybrano wiek respondentów, gdyż jest on silnie skorelowany ze stażem pracy ($R = 0,973$) oraz lepiej opisuje poziom płac [Kot 1999]. Dokładną specyfikację zmiennych przedstawia Tabela 2.

Tabela 2. Zmienne wykorzystane w prezentowanych modelach

w	logarytm naturalny wynagrodzenia progowego
d	logarytm naturalny dochodu
X	wiek liczony w latach
S	stanowisko: asystent - wartość 1, adiunkt - wartość 2, adiunkt z habilitacją - wartość 3, profesor nadzwyczajny - wartość 4, profesor zwyczajny - wartość 5
G	płeć: mężczyzna - wartość 1, gdy kobieta - wartość 0
T	dorabia poza uczelnią macierzystą: gdy tak - wartość 1, nie - wartość 0
K_{ad}	adiunkt - wartość 1, w pozostałych przypadkach - wartość 0
K_{adh}	adiunkt z habilitacją - wartość 1, w pozostałych przypadkach - wartość 0

K_{pnzw}	profesor nadzwyczajny - wartość 1, w pozostałych przypadkach - wartość 0
K_{pzw}	profesor zwyczajny - wartość 1, w pozostałych przypadkach - wartość 0

Źródło: opracowanie własne

WYNIKI

W pierwszej kolejności wyznaczone zostały cztery modele płac typu Mincera, dwa dla płacy progowej (model 1 i model 2) oraz dwa dla dochodu (model 3 oraz model 4). Model 1 oraz model 3 są modelami liniowej zależności zaś w modelach 2 i 4 wykorzystano szeroko opisaną w literaturze zależność wielomianową. Dokładne postacie modeli przedstawione są poniżej.

Model 1

$$\ln(w_i) = a_0 + a_1 \cdot T + a_2 \cdot X + a_3 \cdot G + a_4 \cdot K_{ad} + a_5 \cdot K_{adh} + a_6 \cdot K_{pnzw} + a_7 \cdot K_{pzw} + \varepsilon_i$$

Model 2

$$\ln(w_i) = a_0 + a_1 \cdot T + a_2 \cdot X + a_3 \cdot X^2 + a_4 \cdot X^3 + a_5 \cdot G + a_6 \cdot S + a_7 \cdot S^2 + \varepsilon_i$$

Model 3

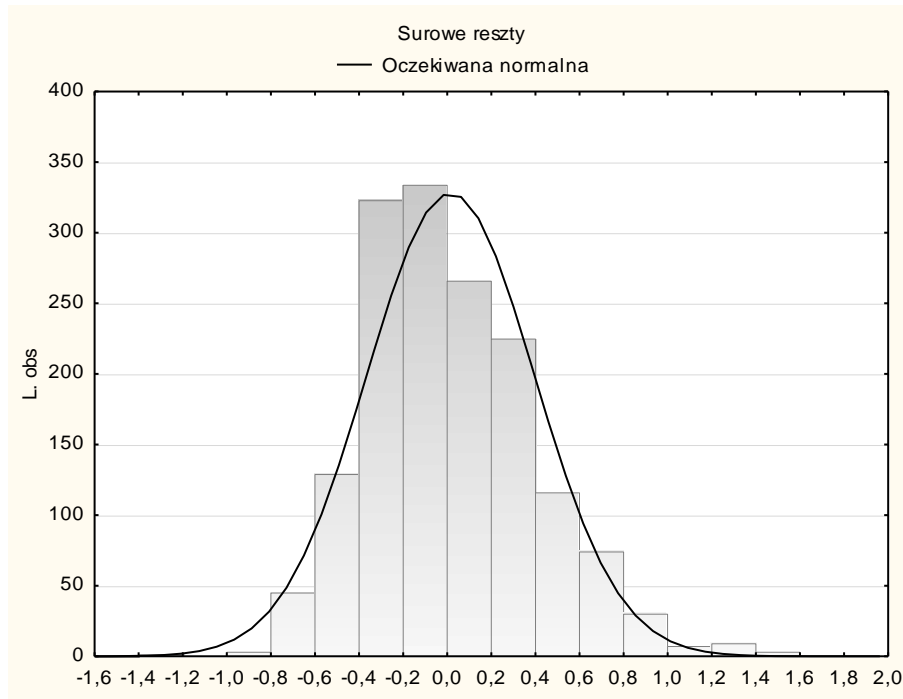
$$\ln(d_i) = a_0 + a_1 \cdot T + a_2 \cdot X + a_3 \cdot G + a_4 \cdot K_{ad} + a_5 \cdot K_{adh} + a_6 \cdot K_{pnzw} + a_7 \cdot K_{pzw} + \varepsilon_i$$

Model 4

$$\ln(d_i) = a_0 + a_1 \cdot T + a_2 \cdot X + a_3 \cdot X^2 + a_4 \cdot X^3 + a_5 \cdot G + a_6 \cdot S + a_7 \cdot S^2 + \varepsilon_i$$

Jedno z podstawowych założeń zastosowanej metody wyznaczania i oceny współczynników modeli wymagało rozkładu normalnego otrzymanych reszt. Z powodów wcześniej opisanych dokonano to poprzez ocenę histogramów reszt. We wszystkich czterech przypadkach nie było rażących odstępstw od oczekiwanej normalnej. Przykładem takiej oceny jest Rysunek 1.

Rysunek 1. Surowe reszty na tle oczekiwanej krzywej normalnej dla modelu 1



Źródło: opracowanie własne przy użyciu programu Statistica 10

Kolejne założenie o homoskedastyczności zweryfikowano testem White'a przyjmując poziom istotności $\alpha = 0,05$. Ponieważ dla modelu 3 i 4 test ten nakazywał odrzucenie hipotezy zerowej do oceny błędów standardowych użyto estymatora Huber-White'a zwanego potocznie „kanapkowym”. Wyniki oszacowań dla wyznaczonych modeli przedstawiają poniższe tabele.

Tabela 3. Oceny parametrów modelu 1

$\ln(w_i)$	parametr	błąd standardowy	z	$P > z $
T	0,1600184	0,019457	8,22	0,000
X	0,0048445	0,001379	3,51	0,000
G	0,1060614	0,0198689	5,34	0,000
K_{adh}	0,3916986	0,0310213	12,63	0,000
K_{adh}	0,6916645	0,0582922	11,87	0,000
K_{pzw}	0,8188634	0,0473392	17,30	0,000
K_{pzw}	0,9617728	0,0590833	16,28	0,000
stała	8,101558	0,0517887	156,43	0,000
$R^2 = 0,4542$ $RMSE = 0,37729$ $F(7,1557) = 185,11$ ($p = 0,0000$)				

Źródło: opracowanie własne z użyciem pakietu Stata 12

Tablica 4. Oceny parametrów modelu 2

$\ln(w_i)$	parametr	błąd standardowy	z	$P > z $
T	0,1559131	0,0193785	8,05	0,000
X	0,0845319	0,0256975	3,29	0,001
X^2	-0,0014442	0,0005079	-2,84	0,005
X^3	0,00000813	0,00000327	2,48	0,013
G	0,108479	0,019804	5,48	0,000
S	0,4125297	0,0560723	7,36	0,000
S^2	-0,0306066	0,008951	-3,42	0,001
stała	6,415316	0,3912656	16,40	0,000
$R^2 = 0,4574$ $RMSE = 0,37619$ $F(7,1557) = 187,52$ ($p = 0,0000$)				

Źródło: opracowanie własne z użyciem pakietu Stata 12

Tablica 5. Oceny parametrów modelu 3

$\ln(d_i)$	parametr	błąd standardowy	z	$P > z $
T	0,3827293	0,024501	15,62	0,000
X	0,0051465	0,0018615	2,76	0,006
G	0,1009102	0,0259889	3,88	0,000
K_{adh}	0,3429024	0,0448158	7,65	0,000
K_{adh}	0,4176427	0,0712181	5,86	0,000
K_{pzw}	0,7376633	0,0663263	11,12	0,000
K_{pzw}	0,8900061	0,0745313	11,94	0,000
stała	7,957867	0,071587	111,16	0,000
$R^2 = 0,3870$ $RMSE = 0,45623$ $F(7,1557) = 114,81$ ($p = 0,0000$)				

Źródło: opracowanie własne z użyciem pakietu Stata 12

Tablica 6. Oceny parametrów modelu 4

$\ln(d_i)$	parametr	błąd standardowy	z	$P > z $
T	0,3839765	0,0245306	15,65	0,000
X	0,0890766	0,0277757	3,21	0,001
X^2	-0,0018309	0,0005102	-3,59	0,000
X^3	0,0000127	0,00000299	4,25	0,000
G	0,1002058	0,025816	3,88	0,000
S	0,3626509	0,0754832	4,80	0,000
S^2	-0,0247951	0,0117828	-2,10	0,036
stała	6,430035	0,4391086	14,64	0,000
$R^2 = 0,3879$ $RMSE = 0,45590$ $F(7,1557) = 127,91$ ($p = 0,0000$)				

Źródło: opracowanie własne z użyciem pakietu Stata 12

Na podstawie wyników estymacji wszystkich modeli można sądzić, że wynagrodzenie progowe jak i dochód w przypadku grupy mężczyzn jest przeciętnie o około 10% wyższe niż dla grupy kobiet. Osoby, które posiadają dodatkowy dochód poza wynagrodzeniem na uczelni macierzystej byłoby skłonni zrezygnować z „dorabiania” podając wysokość płacy progowej średnio o 16% wyższą niż pozostałe osoby. Dodatkowo ta sama grupa pracowników naukowo-

dydaktycznych posiada o około 38% wyższy średni miesięczny dochód brutto niż grupa „niedorabiających” poza uczelnią. Patrząc na zajmowane stanowiska widzimy, że adiunkci, adiunkci z habilitacją, profesorowie nadzwyczajni i profesorowie zwyczajni akceptują hipotetyczny kontrakt dla wynagrodzenia o odpowiednio 39%, 69%, 82% i 96% wyższego od tego co przyjmują asystenci, zaś łączny dochód jest odpowiednio wyższy o 34%, 42%, 74% i 89% od dochodu grupy asystentów. Widać wyraźnie, że największego wzrostu wynagrodzenia w stosunku do faktycznego dochodu oczekuje grupa adiunktów z habilitacją.

Ciekawych spostrzeżeń daje również obraz profili wiekowo-płacowych dla modelu 2 oraz modelu 4. W celu wyznaczenia zależności płacy progowej i dochodu od wieku pracownika w miejsca pozostałych zmiennych objaśniających podstawione zostały wartości przeciętne, które przedstawia tabela 7.

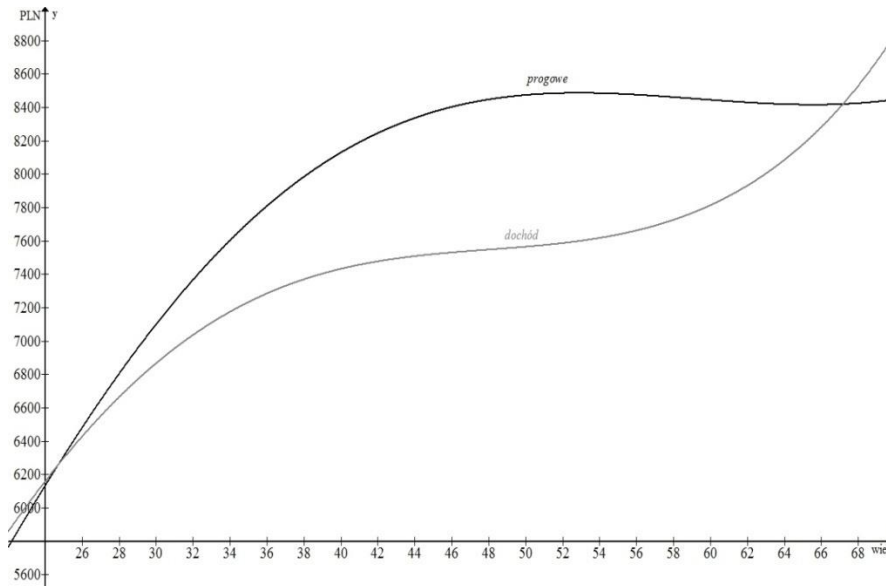
Tabela 7. Wartości przeciętnych wybranych zmiennych

zmienna	średnia
w	8,940
d	8,875
X	42,896
S	2,508
G	0,587
T	0,585
K_{ad}	0,561
K_{adh}	0,042
K_{pnzw}	0,179
K_{pzw}	0,082

Źródło: opracowanie własne z użyciem Statistica 10

Uzyskane zależności pozwalają stwierdzić, że płaca progowa dominuje nad faktycznym, łącznym dochodem w przedziale wieku od 25 do 67 lat. Powyżej 67 lat dochód wyraźnie przewyższa wynagrodzenie progowe. Takiego stanu rzeczy można dopatrywać się w tym, iż wielu pracowników mimo przejścia na emeryturę wciąż pracuje zatem posiada wysokie dochody, zaś ich potrzeby a zarazem oczekiwania płacowe nie rosną z wiekiem. Nadto funkcja wynagrodzenia progowego posiada maksimum lokalne dla wieku 53 lat zaś funkcja dochodu nie posiada żadnego ekstremum lokalnego. Wnioski te widoczne są na poniższym rysunku.

Rysunek 2. Profile wiekowo-płacowe dla płacy progowej i dochodu



Źródło: opracowanie własne z użyciem programu Graph 4.4.

PODSUMOWANIE

W prezentowanej analizie modele regresyjne płac typu Mincera ukazywały zależność logarytmu płacy progowej i dochodu pracowników naukowo-dydaktycznych w zależności od wieku, zajmowanego stanowiska oraz płci. Niezależnie od zastosowanej postaci modelu wnioski były podobne. Kobiety otrzymują mniejsze wynagrodzenia i mniejsze też mają oczekiwania. Im wyższe stanowisko tym wyższy dochód i płaca progowa z tym, że oczekiwania grupy adiunktów z habilitacją są relatywnie wyższe niż pozostałych. Podobnie, wprowadzając warunek jednoetatowości, większego wynagrodzenia oczekują osoby mające dodatkowy dochód poza uczelnią niż osoby niedorabiające. Profile płacowo-wiekowe pokazują natomiast, że wraz z wiekiem dochód typowego uczonego rośnie zaś płaca progowa osiąga swoje maksimum dla 53 lat i wynosi niespełna 8500 zł brutto.

BIBLIOGRAFIA

- Becker G. (1962) Human Capital: A Theoretical and Empirical Analysis, with Special References to Education, NY, Columbia University Press.
- Becker G. (1962) Investment in Human Capital: A Theoretical Analysis, Journal of Political Economy, Vol. 70, pp. 9-49.

- Behrman J., Birdsall N. (1983) The Quality of Schooling: Quantity Alone Is Misleading, *American Economic Review*, Vol. 73, No. 5, 1983, pp. 928-946.
- Ben-Porath Y. (1967) The Production of Human Capital and Life Cycle of Earnings, *The Journal of Political Economy*, Vol. 75, No. 4, p. 352-365.
- Bierens H. J. (1997) Integrated Conditional Moment Estimation and Testing of Median Regression Models, with an Application to a Mincer-Type Equation, *Econometric Society European Meeting*, Toulouse, 27-30 sierpnia.
- Card D., Krueger A. (1992) Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States, *Journal of Political Economy* 100, February, pp. 1-40.
- Deschênes O. (2001) Unobserved Ability, Comparative Advantage and the Rising Return to Education in the United States: A Cohort-Based Approach, *Princeton University Industrial Relations Section Working Paper No. 465*, August.
- Driffill J. (1980) Life-Cycles with Terminal Retirement, *International Economic Review*, Vol. 21, No. 1, 1980, pp. 45-62.
- Frazis H., Loewenstein M. (2005) Reexamining the Returns to Training: Functional Form, Magnitude, and Interpretation, *The Journal of Human Resources*, Vol. 40, No. 2, pp. 453-476.
- Greene W. H. (2002) *Econometric analysis*, Fifth edition, NY University, New Jersey.
- GUS (2011) *Szkoły wyższe i ich finance w 2010 r.*, GUS, Warszawa.
- Haley W. (1973) Human Capital: The Choice between Investment and Income, *American Economic Review*, Vol. 63, No. 5, 1973, pp. 929-944.
- Haley W. (1976) Estimation of the Earnings Profile from Optimal Human Capital Accumulation, *Econometrica* Vol. 44, No. 6, 1976, pp. 1223-1238.
- Hamlen S., Hamlen W. (2012) The Inconsistency of the Quadratic Mincer Equation: A Proof, *Theoretical Economics Letters*, no. 2, p. 115-120.
- Heckman J., Lochner L., Todd P. (2003) Fifty Years of Mincer Earnings Regressions, *NBER WP 9732*.
- Heckman J., Lochner L., Todd P. (2006) Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond, In: E. Hanishek and F. Welch, Eds., *Handbook of the Economics of Education*, Elsevier, Amsterdam, pp. 307-458.
- Heckman J., Polachek S. (1974) Empirical Evidence of the Functional Form of the Earnings-Schooling Relationship, *Journal of the American Statistical Association* Vol. 69, No. 346, pp. 350-354.
- Heckman J., Layne-Farrar A., Todd P. E. (1996) Human Capital Pricing Equations with an Application to Estimating the Effect of Schooling Quality on Earnings, *Review of Economics and Statistics* 78, November, pp. 562-610.
- Kot S. M. (red.) (1999) *Analiza ekonomiczna kształtowania się płac w Polsce w okresie transformacji*, PWN, Warszawa, Kraków.
- Leibowitz A., (1976) Years of Intensity of Schooling Investment, *American Economic Review*, Vol. 66, No. 3, pp. 321-334.
- Lemieux T. (2003) *The Mincer Equation Thirty Years after Schooling, Experience and Earnings*, Center for Labor Economics, University of California-Berkeley, Berkeley.

- Mincer J. (1958) Investment in Human Capital and Personal Income Distribution, *Journal of Political Economy*, Vol. 66, pp. 281-302.
- Mincer, J. (1974) *Schooling, Experience, and Earnings*, Columbia Univ. Press, New York.
- Mincer, J. (1997) Changes in Wage Inequality, 1970-1990, *Research in Labor Economics* 16, pp. 1-18.
- Murphy K., Welch F. (1990) Empirical Age-Earnings Pro-files, *Journal of Labor Economics*, Vol. 8, No. 2, pp. 202-229.
- Ryder H., Stafford F., Stephan P. (1976) Labor, Leisure and Training over the Life-Cycle, *International Economic Review*, Vol. 17, No. 3, p. 651-674.
- Sadeghi J. M., Shirouyehzad L., Samadi S. (2012) Estimating the Impact of Education on Income with Econometric Approach: A Case Study in Universities, *International Journal of Academic Research in Business and Social Sciences*, Vol. 2, No. 6, pp. 175-185.
- Schultz T. (1961) Investment in Human Capital, *American Economic Review*, Vol. 51, pp. 1-17.
- Sheshinski E. (1968) On the Individual's Lifetime Allocation Between Education and Work, *Metroeconomica*, Vol. 20, No. 1, 1968, pp. 42-49.

MINCER-TYPE WAGE EQUATION MODELS IN THE CASE OF POLISH ACADEMICS IN PUBLIC UNIVERSITIES

Abstract: The article presents the results of research which main aim was to determine and compare the Mincer-wage type models for group of polish academics in public universities. The data used for the analysis came from a national survey carried out in the spring of 2012. Compared models were related to salary actually received and the reservation wage for the hypothetical contract with no extra work condition.

Keywords: Mincer-type wage equation, reservation wage, academics salary

DEKOMPOZYCJA RÓŻNIC POMIĘDZY KOBIECIAMI I MĘŻCZYZNAМИ W PROCESIE OPUSZCZANIA STANU BEZROBOCIA

Joanna Małgorzata Landmesser

Katedra Ekonometrii i Statystyki

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

e-mail: joanna_landmesser@sggw.pl

Streszczenie: W pracy przeprowadzono analizę stóp wyjścia ze stanu bezrobocia, uwzględniającą zróżnicowanie płciowe. Proces opuszczania stanu bezrobocia badano dla obu płci osobno wykorzystując narzędzia z zakresu analizy czasu trwania – parametryczne modele hazardu. Głównym celem było dokonanie rozkładu nierówności między kobietami i mężczyznami podczas opuszczania bezrobocia. Zastosowana zmodyfikowana mikroekonometryczna technika dekompozycji Oaxaca-Blindera pozwoliła na wyodrębnienie czynników wyjaśniających zaobserwowane nierówności.

Słowa kluczowe: czas trwania w bezrobociu, nierówności płciowe, parametryczne modele hazardu, dekompozycja Oaxaca-Blindera

WSTĘP

Zachowanie kobiet i mężczyzn na rynku pracy jest zróżnicowane. Tradycyjnie mężczyźni są silniej związani z tym rynkiem, kobiety zaś – z uwagi na częstsze poświęcanie się karierze rodzinnej – w słabszym stopniu. Postawy te znajdują odzwierciedlenie w szansach na podejmowanie zatrudnienia.

Literatura zawierająca wyniki badań nad zróżnicowanym poziomem aktywności ekonomicznej kobiet i mężczyzn jest obszerna (jej przeglądu dostarcza np. [Altonji, Blank 1999]). Badania skoncentrowane na różnicach w stopach wyjścia ze stanu bezrobocia między kobietami i mężczyznami są znacznie rzadsze. Płeć jest ważnym czynnikiem wpływającym na przepływy między stanami na rynku pracy. Studia pokazują, że kobiety częściej niż mężczyźni narażone są na

okresy bez pracy (np. [Steiner 1989]), natomiast w [Katz, Meyer 1990] stwierdzono, że cechuje je niższe prawdopodobieństwo znalezienia zatrudnienia.

Analizy skupione na opisie wyjść ze stanu bezrobocia, które dokonano w Polsce do tej pory, traktują kwestię płci w sposób niezbyt szczegółowy. Większość z nich pomija to, że wpływ determinant na szansę podjęcia pracy może się różnić w zależności od płci. W budowanych modelach ekonometrycznych (np.: [Małarska 2007]) najczęściej uwzględniana jest dychotomiczna zmienna „płeć” (co skutkuje odpowiednim przesunięciem wartości zmiennej objaśnianej), natomiast nie praktykuje się estymacji osobnych równań dla kobiet i mężczyzn. Odmienne podejście proponują za granicą [Gonzalo, Saarela 2000] czy [Tansel, Taşçı 2010], a w Polsce [Landmesser 2008], którzy szacują oddzielne modele hazardu dla obu płci.

Przeprowadzone badanie stanowi wnikliwszą analizę stóp wyjścia ze stanu bezrobocia, uwzględniającą zróżnicowanie płciowe. W artykule proces opuszczania stanu bezrobocia zostanie zbadany dla kobiet i mężczyzn osobno, wykorzystując narzędzia z zakresu analizy czasu trwania - parametryczne modele hazardu. Głównym celem badania jest dokonanie rozkładu obserwowanych nierówności między kobietami i mężczyznami w trakcie opuszczania stanu bezrobocia. Zmodyfikowana technika dekompozycji Oaxaca-Blindera pozwoli na wyodrębnienie czynników wyjaśniających nierówności.

METODA ANALIZY

Badaniom poddano czas trwania przez osobę w stanie bezrobocia (T). Modelowanie tej zmiennej wymaga zastosowania narzędzi z zakresu analizy przeżycia. Konstruowane modele hazardu pozwalają uwzględnić obserwacje cenzurowane oraz wpływ cech indywidualnych jednostek na szanse opuszczenia wyjściowego stanu bezrobocia (por. [Lancaster 1979]).

Podstawową funkcją opisującą rozkład czasów trwania jest funkcja przeżycia $S(t) = \Pr[T > t] = 1 - F(t)$, wyrażająca prawdopodobieństwo przeżycia chwili czasu t (por. [Kalbfleisch, Prentice 2002]). Iloraz funkcji gęstości oraz funkcji przeżycia definiuje funkcję hazardu (ryzyka, intensywności). Określa ona bezpośrednie ryzyko tego, że pewne zdarzenie nastąpi w przedziale czasowym pomiędzy t i $t+dt$, pod warunkiem, że do momentu t zajście jego nie nastąpiło, co wyraża następujący wzór:

$$h(t) = \frac{f(t)}{S(t)} = \lim_{dt \rightarrow 0} \frac{\Pr[t \leq T < t + dt \mid T \geq t]}{dt} \quad (1)$$

Za pomocą parametrycznych modeli proporcjonalnych hazardów możliwe jest wyspecyfikowanie hazardu jako funkcji czasu oraz zmiennych objaśniających:

$$h(t \mid \mathbf{x}_j) = h_0(t) \exp(\mathbf{x}'_j \boldsymbol{\beta}) \quad (2)$$

Do opisu monotonicznego hazardu często wykorzystywany jest dwuparametryczny rozkład Weibulla $W(\lambda, p)$. Weibull opisał nim dyspersję trwałości zmęczeniowej materiałów oraz dyspersję czasów pracy maszyn i urządzeń [Weibull 1939], ale jest on również stosowany przez badaczy modelujących czas trwania w bezrobociu (ze względu na malejącą w czasie intensywność opuszczania tego stanu). Dla zmiennej $T \sim W(\lambda, p)$ funkcje przeżycia i hazardu zdefiniowane są jako $S(t) = \exp(-\lambda t^p)$ oraz $h(t) = p t^{p-1} \lambda$, dla $\lambda > 0, p > 0$ oraz $t \geq 0$. Estymowany na podstawie danych empirycznych parametr p jest tzw. parametrem kształtu. Jeśli $p = 1$, to funkcja hazardu jest stała; jeśli $p > 1$, to jest monotonicznie rosnąca; natomiast dla $p < 1$ hazard monotonicznie maleje. Postać modelu hazardu Weibulla po podjęciu parametryzacji $\lambda = \exp(\mathbf{x}'\boldsymbol{\beta})$ jest następująca:

$$h(t|\mathbf{x}_j) = p t^{p-1} \exp(\mathbf{x}'_j \boldsymbol{\beta}) \quad (3)$$

Celem opisu procesu opuszczania stanu bezrobocia przez kobiety i mężczyzn w pracy szacowano parametryczne modele proporcjonalnych hazardów Weibulla dla obu płci osobno. Następnie dokonano rozkładu zaobserwowanych nierówności stosując modyfikację techniki rozkładu różnic Oaxaca-Blindera.

Przyjmując oznaczenia:

$\mathbf{x}^K, \mathbf{x}^M$ - wektory charakterystyk kobiet i mężczyzn,

$\boldsymbol{\beta}^K, \boldsymbol{\beta}^M$ - parametry powiązane z charakterystykami osób,

$\bar{\Phi}$ - przeciętny poziom zmiennej objaśnianej w modelu ekonometrycznym (np. średni poziom hazardu),

popularna dekompozycja Oaxaca-Blindera ([Oaxaca 1973], [Blinder 1973]) przedstawia się następująco:

$$\bar{\Phi}(\mathbf{x}^K \boldsymbol{\beta}^K) - \bar{\Phi}(\mathbf{x}^M \boldsymbol{\beta}^M) = [\bar{\Phi}(\mathbf{x}^K \boldsymbol{\beta}^M) - \bar{\Phi}(\mathbf{x}^M \boldsymbol{\beta}^M)] + [\bar{\Phi}(\mathbf{x}^K \boldsymbol{\beta}^K) - \bar{\Phi}(\mathbf{x}^K \boldsymbol{\beta}^M)] \quad (4)$$

Pierwszy wyraz po prawej stronie równania (4) wyraża różnicę tzw. potencjałów dwóch grup osób (kobiet i mężczyzn). Ten składnik odpowiada części zróżnicowania cechy zależnej, która wynika z faktu, iż kobiety i mężczyźni są „różni” i mogą wykonywać „różną” pracę. Gdyby charakterystyki \mathbf{x} użyte do estymacji modelu ekonometrycznego wyczerpywały wszystkie czynniki wpływające na szanse opuszczenia stanu bezrobocia, to można by przyjąć, że drugi wyraz prawej strony powyższego równania reprezentuje wielkość dyskryminacji. Wyraz ten to efekt różnic w wyestymowanych parametrach równań, a więc „cenach” poszczególnych cech mężczyzn i kobiet. Blinder twierdził, że „występuje [on] tylko wtedy, gdy rynek różnie wynagradza identyczne zestawy cech, jeśli należą do członków innych [...] grup” [Blinder 1973].

Modyfikacja metody Oaxaca-Blindera dokonana przez Yuna pozwoliła na wykorzystanie powyższej koncepcji do określenia wkładu poszczególnych zmiennych objaśniających z modelu regresji Weibulla w wyjaśnianiu nierówności

międzyplciowych [Yun 2004]. Wzór podany przez Yuna ma następującą postać (za [Ortega Masagué 2008]):

$$\begin{aligned} & \overline{\Phi}(\mathbf{x}^K \boldsymbol{\beta}^K) - \overline{\Phi}(\mathbf{x}^M \boldsymbol{\beta}^M) = \\ & = \sum_{i=1}^k W_{\Delta x_i} [\overline{\Phi}(\mathbf{x}^K \boldsymbol{\beta}^M) - \overline{\Phi}(\mathbf{x}^M \boldsymbol{\beta}^M)] + \sum_{i=1}^k W_{\Delta \beta_i} [\overline{\Phi}(\mathbf{x}^K \boldsymbol{\beta}^K) - \overline{\Phi}(\mathbf{x}^K \boldsymbol{\beta}^M)], \end{aligned} \quad (5)$$

gdzie:

$$\begin{aligned} W_{\Delta x_i} &= \frac{(\bar{x}_i^K - \bar{x}_i^M) \beta_i^M}{(\bar{\mathbf{x}}^K - \bar{\mathbf{x}}^M)' \boldsymbol{\beta}^M}, & \sum_{i=1}^k W_{\Delta x_i} &= 1, \\ W_{\Delta \beta_i} &= \frac{\bar{x}_i^K (\beta_i^K - \beta_i^M)}{\bar{\mathbf{x}}^K (\boldsymbol{\beta}^K - \boldsymbol{\beta}^M)}, & \sum_{i=1}^k W_{\Delta \beta_i} &= 1, \end{aligned} \quad (6)$$

natomiast k oznacza liczbę zmiennych objaśniających w modelu, a $\bar{\mathbf{x}}^K, \bar{\mathbf{x}}^M$ - średnie poziomy charakterystyk dla kobiet i mężczyzn odpowiednio.

OPIS DANYCH EMPIRYCZNYCH

Badanie przeprowadzono wykorzystując dane indywidualne osób zarejestrowanych jako bezrobotne w Powiatowym Urzędzie Pracy w Słupsku. Wylosowana próba obejmowała 4372 osoby, zarejestrowane w PUP od stycznia 1990 do sierpnia 2007. Wylosowano 2203 kobiety oraz 2169 mężczyzn (kobiety stanowiły 50,4%, a mężczyźni 49,6% wylosowanej próby). Dane o każdej osobie miały postać wieloepizodową i zawierały szczegółowe zapisy historii klientów urzędu. Na ich podstawie ustalono, jak długo trwały (w dniach) poszczególne epizody bycia bezrobotnym lub jak długo wciąż jeszcze one trwają (w wypadku epizodów cenzurowanych). Każda osoba w swojej historii mogła wykazywać wielokrotne rejestracje w PUP, stąd dla 4372 analizowanych osób zanotowano w sumie 10118 epizodów bycia bezrobotnym, z czego 10 % stanowiły epizody cenzurowane (Tabela 1).

Tabela 1. Liczebność epizodów bezrobocia

Liczba epizodów	Razem		Kobiety		Mężczyźni	
ogółem	10118	100%	4786	100%	5332	100%
cenzurowanych	1007	10%	664	13,9%	343	6,4%
niecenzurowanych	9111	90%	4122	86,1%	4989	93,6%

Źródło: obliczenia własne na podstawie danych z PUP w Słupsku

W przypadku kobiet epizody cenzurowane stanowiły większy udział wśród wszystkich epizodów niż miało to miejsce dla mężczyzn. Wiązało się to zapewne z faktem, że kobiety doznawały przeciętnie dłuższych epizodów bezrobocia niż mężczyźni (przykładowo epizody niecenzurowane u kobiet trwały średnio 413,1 dni, natomiast u mężczyzn 271,2 dni). Informacje na temat średniego czasu trwania w stanie bezrobocia w trakcie pojedynczego epizodu przedstawia Tabela 2.

Tabela 2. Średni czas trwania w stanie bezrobocia

Średnia długość (w dniach)	Razem	Kobiety	Mężczyźni
epizodu bezrobocia	406,5	523,2	301,7
w tym: cenzurowanego	1049	1207	745,5
niecenzurowanego	335,4	413,1	271,2

Źródło: obliczenia własne na podstawie danych z PUP w Słupsku

Przeciętnie pojedyncza kobieta rejestrowała się w urzędzie pracy 2,55 razy, natomiast mężczyzna 3,02 razy (częściej, ale za to na krócej). Średni wiek kobiety w chwili rozpoczęcia epizodu bezrobocia wyniósł 31,99 lat, z kolei dla mężczyzny było to 33,66 lat. Szczegółowe informacje na temat wieku rejestrujących się osób, poziomu ich wykształcenia oraz miejsca zamieszkania przedstawia Tabela 3.

Tabela 3. Struktura epizodów bezrobocia według wybranych cech

Cecha		Razem		Kobiety		Mężczyźni	
wiek	od 17 do 24 lat	3068	30,3%	1512	31,6%	1556	29,2%
	od 25 do 34 lat	2755	27,2%	1353	28,3%	1402	26,3%
	od 35 do 44 lat	2408	23,8%	1200	25,1%	1208	22,7%
	od 45 do 54 lat	1742	17,2%	699	14,6%	1043	19,6%
	powyżej 55 lat	145	1,4%	22	0,5%	123	2,3%
wykształcenie	wyższe	767	7,6%	517	10,8%	250	4,7%
	średnie zaw.	1957	19,3%	1170	24,4%	787	14,8%
	średnie ogóln.	710	7%	516	10,8%	194	3,6%
	zawodowe	3100	30,6%	1240	25,9%	1860	34,9%
	gimnazjalne	3584	35,4%	1343	28,1%	2241	42%
miejsce zam.	miasto	5079	50,2%	2551	53,3%	2528	47,4%
	wieś	5039	49,8%	2235	46,7%	2804	52,6%

Źródło: obliczenia własne na podstawie danych z PUP w Słupsku

Rejestrujące się w urzędzie pracy kobiety były zazwyczaj młodsze od mężczyzn. Częściej niż mężczyźni posiadały wyższe wykształcenie lub średnie, rzadziej zawodowe i niższe. Charakteryzuje je także wyższy odsetek zamieszkania w mieście. Epizody bezrobocia zbadano także pod kątem wypłaty w trakcie ich trwania zasiłku, dodatku szkoleniowego oraz świadczenia z ZUS. Okazało się, że mężczyźni częściej niż kobiety byli beneficjentami zasiłków dla bezrobotnych, kobietom zaś częściej przydzielano dodatki szkoleniowe oraz świadczenia ZUS.

Informacje uzyskane z bazy danych z PUP pozwoliły na wyłonienie zestawu potencjalnych zmiennych objaśniających w modelach opisujących intensywność opuszczania stanu bezrobocia. Większość z nich to zmienne dychotomiczne, np.: „płeć” (cyfrą 1 zakodowano płeć męską), zestaw zmiennych dla pięciu kategorii wiekowych, zestaw zmiennych dotyczących poziomu wykształcenia osoby, „związek” (1, jeśli osoba nie jest wolnego stanu cywilnego), „miasto” (1, jeśli

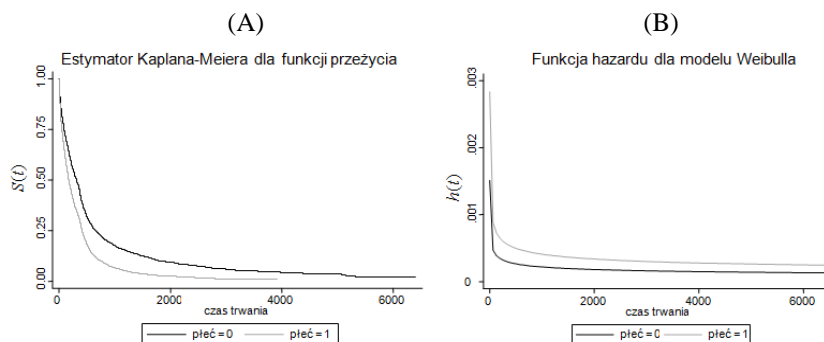
osoba mieszka w mieście), „niesprawność” (1, jeśli osoba jest niepełnosprawna), „zasiłek” (1 dla beneficjentów zasiłku dla bezrobotnych), „dodatek szkol” (1, dla otrzymujących dodatek szkoleniowy), „zus” (1 dla świadczeniobiorców ZUS). Ponadto w estymowanych modelach wykorzystano informację o numerze kolejnego epizodu bezrobocia dla osoby (zmienna „nrepiz”) oraz o poziomie stopy bezrobocia w miesiącu rejestracji („stopabezrob”).

Zależne od czasu zmienne objaśniające („zasiłek”, „dodatek szkol”, czy „zus”) zostały uwzględnione w modelach poprzez zastosowanie metody rozszczepienia epizodów. Zgodnie z ideą tej metody dokonano podziału oryginalnych epizodów na subepizody w tych punktach czasu, w których miała miejsce zmiana wartości zmiennej zależnej od czasu.

PREZENTACJA UZYSKANYCH WYNIKÓW

Przystępując do analizy rozkładu czasów trwania w bezrobociu, w pierwszej kolejności sporządzono wykres nieparametrycznego estymatora Kaplana-Meiera dla funkcji przeżycia $S(t)$, osobno dla kobiet i mężczyzn (Rysunek 1 (A)).

Rysunek 1. Wykresy funkcji przeżycia $S(t)$ oraz funkcji hazardu $h(t)$ względem płci



Źródło: obliczenia własne z wykorzystaniem programu Stata

Niższe położenie krzywej dla mężczyzn wskazuje na mniejsze prawdopodobieństwo dalszego trwania w stanie bezrobocia dla tej płci. Zaobserwowano statystyczną istotność różnic dla rozważanych funkcji przeżycia.

Następnie na podstawie całej próby oszacowano parametryczny model proporcjonalnych hazardów Weibulla dla szansy opuszczenia stanu bezrobocia (oceny parametrów tego modelu zawiera Tabela 4, część (A)). W modelu tym przykładowa interpretacja dla oceny parametru stojącego przy zmiennej „płeć” jest następująca: szansa na opuszczenie stanu bezrobocia przez mężczyznę jest o 87,5% większa niż przez kobietę ($\exp(0,629) = 1,875$). Dodatnia wartość parametru β_k oznacza, że zmiana wartości k -tej zmiennej objaśniającej o jednostkę wywoła dodatni względny przyrost stopy hazardu (odpowiednio ujemna ocena wiąże się ze

względny spadek hazardu). Wnioskujemy, że szanse na opuszczenie stanu bezrobocia wzrastają w przypadku płci męskiej, młodszego wieku, lepszego wykształcenia, trwania w związku, mieszkania w mieście, kolejnego epizodu bezrobocia w karierze, natomiast szanse te umniejsza niepełnosprawność, wyższa stopa bezrobocia w chwili rejestracji w urzędzie pracy oraz fakt otrzymywania zasiłku dla bezrobotnych, dodatku szkoleniowego oraz świadczenia z ZUS. Rysunek 1 (B) przedstawia dwa wykresy funkcji hazardu Weibulla otrzymane na podstawie oszacowanego modelu, przy założeniach „płeć” = 0 oraz „płeć” = 1. Funkcje te są malejące (ponieważ $p < 0$); jedna stanowi przesunięcie drugiej; w każdej chwili niższa intensywność opuszczania bezrobocia charakteryzuje kobiety.

Tabela 4. Wyniki estymacji modeli Weibulla dla całej próby (A), osobno dla kobiet (B) oraz osobno dla mężczyzn (C)

Zmienna	Razem (A)		Kobiety (B)	Mężczyźni (C)
	β_i^R	$\exp(\beta_i^R)$	β_i^K	β_i^M
płeć	0,629 ***	1,875	–	–
wiek17-24	0,570 ***	1,769	0,943 ***	0,774 ***
wiek25-34	0,584 ***	1,793	1,137 ***	0,632 ***
wiek35-44	0,442 ***	1,556	1,125 ***	0,348 ***
wiek45-54	0,365 ***	1,441	1,102 ***	0,258 **
wyższe	0,682 ***	1,977	0,979 ***	0,368 ***
średnie	0,482 ***	1,619	0,601 ***	0,388 ***
średnie	0,500 ***	1,648	0,570 ***	0,483 ***
nizsze	0,264 ***	1,301	0,384 ***	0,210 ***
wzrost	0,239 ***	1,270	-0,055	0,491 ***
miasto	0,058 ***	1,060	0,064 **	0,056 *
niepełnosprawność	-0,385 ***	0,680	-0,236 ***	-0,431 ***
stopa bezrobocia	-0,046 ***	0,955	-0,040 ***	-0,051 ***
epizod	0,063 ***	1,065	0,099 ***	0,043 ***
zasiłek	-7,490 ***	0,001	-19,599	-7,085 ***
dodatek szkolny	-3,539 ***	0,029	-3,726 ***	-3,324 ***
zUS	-3,600 ***	0,027	-3,474 ***	-4,104 ***
cons	-4,381 ***	0,013	-4,857 ***	-3,923 ***
p	0,721 ***	–	0,697 ***	0,755 ***
liczba epizodów	10118		4786	5332
liczba subepizodów	22373		10890	11483
lnL	-13188,137		-6436,8563	-6579,658

Źródło: obliczenia własne z wykorzystaniem programu Stata;

***, **, * - istotność statystyczna na poziomie 1%, 5%, 10%

W kolejnym kroku oszacowano dwa modele hazardu Weibulla dla czasu spędzonego w stanie bezrobocia, dla kobiet i mężczyzn osobno (wyniki estymacji zawiera Tabela 4, części (B) i (C)). Modele te różnią się ocenami parametrów stojących przy odpowiednich zmiennych. Najwyższą skłonność do opuszczania

stanu bezrobocia posiadają kobiety w wieku od 25 do 34 lat, natomiast w przypadku mężczyzn największe szanse mają ci w wieku od 17 do 24 lat. U kobiet najsilniejszy jest pozytywny efekt wyższego wykształcenia, u mężczyzn zaś efekt wykształcenia średniego ogólnokształcącego. Trwanie w związku małżeńskim intensyfikuje wyjścia z bezrobocia u mężczyzn. Poziom odpowiedniego parametru dla kobiet wskazywałby na spadek szans na podjęcie pracy, parametr ten jest jednak statystycznie nieistotny. Spośród pozostałych parametrów na uwagę zasługują te przy zmiennych dotyczących świadczeń finansowych. O ile negatywny wpływ dodatków szkoleniowych oraz świadczeń z ZUS na szanse kobiet i mężczyzn na opuszczenie bezrobocia jest zbliżony, to nie można stwierdzić tego samego o skutkach zasiłków dla bezrobotnych. Odpowiedni parametr w równaniu modelu hazardu dla kobiet – pomimo, że jest ujemny – okazał się statystycznie nieistotny. Oszacowane wartości parametru p wskazują na gwałtowniejszy spadek szansy na zatrudnienie w miarę upływu czasu w przypadku kobiet.

Otrzymane oceny parametrów w modelach (B) i (C) są jednak ze względu na różny skład oraz różną liczebność obu podprób trudno ze sobą porównywalne. Stąd w dalszej kolejności dokonano rozkładu zaobserwowanych nierówności między kobietami i mężczyznami w trakcie opuszczania stanu bezrobocia stosując zmodyfikowaną technikę dekompozycji różnic Oaxaca-Blindera. Jej wyniki w sposób zagregowany przedstawia Tabela 5 (agregacja polegała na skumulowaniu efektów grup powiązanych zmiennych).

Tabela 5. Wyniki dekompozycji różnic zmodyfikowaną techniką Oaxaca-Blindera

Zaobserwowana różnica			Wartość		%
			-0,0027787		100
	Wartość	%		Wartość	%
<i>Charakterystyki</i>	-0,000085	3,05	<i>Parametry</i>	-0,002694	96,95
wiek	-0,000029	1,03	wiek	0,000708	-25,50
wykształcenie	-0,000087	3,14	wykształcenie	0,000228	-8,22
związek	-0,000051	1,83	związek	-0,000476	17,13
miasto	-0,000005	0,17	miasto	0,000006	-0,21
niesprawność	-0,000005	0,19	niesprawność	0,000010	-0,35
stopabezrob	-0,000007	0,25	stopabezrob	0,000225	-8,08
nrepiz	0,000024	-0,86	nrepiz	0,000178	-6,41
zasiłek	-0,000312	11,23	zasiłek	-0,003634	130,78
dodatekszkol	0,000077	-2,77	dodatekszkol	-0,000027	0,96
zus	0,000310	-11,17	zus	0,000087	-3,14

Źródło: obliczenia własne z wykorzystaniem programu Stata

Ma miejsce ujemna różnica pomiędzy średnimi wartościami funkcji hazardu dla kobiet i mężczyzn, co oznacza, że kobiety mają przeciętnie niższe szanse na opuszczenie stanu bezrobocia niż mężczyźni. W efekcie przeprowadzonej

dekompozycji wyodrębniono czynniki w różnym stopniu wyjaśniające zaobserwowaną nierówność. W tabeli wytłuszczono te czynniki, które w największym stopniu były „winne” zanotowanemu zróżnicowaniu szans.

Okazuje się, że różnice w stopach wyjścia z bezrobocia w znikomym stopniu – tylko w 3,05% – są wyjaśniane przez cechy indywidualne kobiet i mężczyzn (wektory \mathbf{x}^K oraz \mathbf{x}^M). Objaśniana w ten sposób część różnic w szansach na wyjście z bezrobocia bierze się stąd, że kobiety różnią się od mężczyzn pewnymi cechami istotnymi na rynku pracy. Zauważyć można przemawiający na niekorzyść kobiet efekt poziomu wykształcenia. Pomimo, że kobiety są przeciętnie lepiej wykształcone od mężczyzn, to zdobywają one najczęściej wykształcenie humanistyczne, co w rezultacie wpływa na mniejsze prawdopodobieństwo zatrudnienia niż w wypadku technicznie szkolonych mężczyzn. Ponadto, nierówności zwiększa efekt wsparcia zasiłkiem dla bezrobotnych. Zróżnicowanie w stopach wyjścia z bezrobocia jest zaś łagodzone przez częstsze dodatki szkoleniowe i świadczenia ZUS dla kobiet. Co ciekawe, wyższa liczba zasiłków szkoleniowych w wypadku kobiet nie wynika z dyskryminacji płci, lecz z konieczności podejmowania przez nie szkoleń pozwalających uzyskać kwalifikacje zgodne z oczekiwaniami pracodawców.

Analizowane różnice międzypłciowe w znakomitej jednak większości – w 96,95% – powinny zostać przypisane współczynnikom β^K i β^M szacowanych modeli hazardu, a nie zróżnicowaniu charakterystyk jednostek. Okazuje się, że osoby o takich samych cechach, o ile są różnej płci, mają odmienne szanse na wychodzenie ze stanu bezrobocia. To właśnie różna „wycena” cech kobiet i mężczyzn jest główną przyczyną nierówności. Można przypuszczać, że pracodawcy faworyzują mężczyzn, a kobiety są dyskryminowane na rynku pracy. Zróżnicowanie szans wynika głównie z uprzedzeń powiązanych ze stanem cywilnym kobiety oraz ze względu na silniejszy w jej wypadku efekt zasiłkowy.

PODSUMOWANIE

W artykule proces opuszczania stanu bezrobocia badano dla obu płci osobno wykorzystując parametryczne modele hazardu. Uzyskane wyniki potwierdziły, że wpływ zmiennych objaśniających na szansę podjęcia zatrudnienia zależy od płci jednostki. Ponadto w pracy opisano szczegółowo, jakie są składowe zaobserwowanego zróżnicowania międzypłciowego.

Z przeprowadzonej dekompozycji wynika, że różnice w intensywności opuszczania stanu bezrobocia w małym stopniu są wyjaśniane przez charakterystyki indywidualne kobiet i mężczyzn. Mimo to, zauważono działający na niekorzyść kobiet efekt wyższego poziomu wykształcenia. Można przypuszczać, że struktura rozkładu obu płci pomiędzy poszczególne zawody w gospodarce również będzie niekorzystna dla kobiet, ponieważ sfeminizowane zawody nie są zazwyczaj kojarzone z dużą rotacją w zatrudnieniu. Nierówności zmniejsza efekt częstszych dla kobiet dodatków szkoleniowych i świadczeń ZUS.

W o wiele większym jednak zakresie zaobserwowane różnice mogą zostać przypisane „wycenom” cech kobiet i mężczyzn dokonywanym przez rynek. Ta właśnie część różnic, tzw. część niewyjaśniona, utożsamiana może być z efektem dyskryminacji kobiet na rynku pracy i jest najważniejszym czynnikiem wyjaśniającym zróżnicowanie szans na zatrudnienie wśród kobiet i mężczyzn. Wskazuje ona na fakt, iż zaobserwowane nierówności wystąpiłyby zapewne nawet wówczas, gdyby charakterystyki obu płci były identyczne.

Ze względu na słabsze dopasowanie modelu hazardu dla kobiet (nieistotne dwa parametry) wyniki przeprowadzonej dekompozycji mogą budzić jednak pewne wątpliwości. Większość zaobserwowanego zróżnicowania wynika z wartości parametrów, ale nie wszystkie parametry mają oczekiwane własności statystyczne. Dlatego też wydaje się, że dla poprawnego wnioskowania dokonywanego za pomocą przedstawionej w artykule metody kluczowy jest staranny dobór zmiennych objaśniających do modeli hazardu.

BIBLIOGRAFIA

- Altonji J.G., Blank R.M. (1999) Race and Gender in the Labor Market. [w:] O. Ashenfelter, D. Card (red.) Handbook of Labor Economics, Volume III. North-Holland, Amsterdam.
- Blinder A.S. (1973) Wage Discrimination: Reduced Form and Structural Estimates. *Journal of Human Resources*, Vol. 8(4), str. 436 – 455.
- Gonzalo M.T., Saarela J. (2000) Gender Differences in Exit Rates from Unemployment: Evidence from a Local Finnish Labour Market. *Finnish Economic Papers*, 13(2), str. 129 – 139.
- Kalbfleisch J., Prentice R. (2002) *The Statistical Analysis of Failure Time Data*. Second Edition. John Wiley and Sons, New York.
- Katz L.F., Meyer B.D. (1990) Unemployment Insurance, Recall Expectations, and Unemployment Outcomes. *Quarterly Journal of Economics*, 105(4), str. 973 – 1002.
- Lancaster T. (1979) Econometric Methods for the Duration of Unemployment. *Econometrica*, 47, str. 939 – 956.
- Landmesser J. (2008) Analiza aktywności ekonomicznej kobiet wiejskich z wykorzystaniem ekonometrycznych modeli hazardu. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, Zeszyt 1, Tom X, str. 233–239.
- Malarska A. (2007) Diagnostowanie determinantów bezrobocia w Polsce nieklasycznymi metodami statystycznymi. Analiza empiryczna na podstawie danych Bael. Wydawnictwo Biblioteka, Łódź.
- Oaxaca R.L. (1973) Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, Vol. 14, str. 693 – 709.
- Ortega Masagué A. C. (2008) Gender Gaps in Unemployment Rates in Argentina. *Económica*, La Plata, Vol. LIV, nr 1-2, str. 161 – 202.
- Steiner V. (1989) Causes of Recurrent Unemployment - An Empirical Analysis. *Empirica*, 16, str. 53 – 65.

- Tansel A., Taşçi H.M. (2010) Hazard Analysis of Unemployment Duration by Gender in a Developing Country: The Case of Turkey. IZA Discussion Paper, 4844, Bonn.
- Weibull W. (1939) A Statistical Theory of the Strength of Materials. Ingeniörsvetenskapsakademiens Handlingar, Nr 151, Stockholm.
- Yun M. (2004) Decomposing Differences in the First Moment. Economic Letters, Vol. 82, str. 275 – 280.

**DECOMPOSITION OF DIFFERENCES
BETWEEN WOMEN AND MEN
IN THE PROCESS OF LEAVING THE UNEMPLOYMENT**

Abstract: In the paper, we analyse the exit rates from unemployment, taking into account gender differences. The process of leaving the unemployment state was examined for each sex separately using the tools of duration analysis – the parametric hazard models. The main objective was to perform a decomposition of inequalities between men and women when leaving unemployment. Applied the modified Oaxaca-Blinder microeconomic decomposition technique allowed us to isolate the factors explaining the observed inequalities.

Keywords: duration of unemployment, gender inequalities, parametric hazard models, Oaxaca-Blinder decomposition

DOCHODY INDYWIDUALNE A DOCHODY GOSPODARSTW DOMOWYCH W POLSCE W LATACH 2000 – 2010

Piotr Łukasiewicz, Krzysztof Karpio, Arkadiusz Orłowski
Katedra Informatyki
Szkola Główna Gospodarstwa Wiejskiego w Warszawie
e-mails: piotr_lukasiewicz@sggw.pl, krzysztof_karpio@sggw.pl,
arkadiusz_orlowski@sggw.pl

Streszczenie: W pracy przedstawiono wyniki badań dotyczących zależności pomiędzy rozkładami dochodów indywidualnych i dochodów gospodarstw domowych w Polsce w okresie 2000–2010. W USA rozkład dochodów gospodarstw domowych można wyrazić poprzez splot rozkładów dochodów indywidualnych członków rodziny. Jak wynika z wcześniejszych badań, w Polsce nie obserwujemy takiej zależności. W zasadzie, mamy bardzo małą wiedzę o rodzaju stosunków między tymi kategoriami dochodów naszym kraju. Przypuszcza się, że w USA dochody osobiste członków rodzin są zmiennymi niezależnymi. Z drugiej strony, w Polsce są na pewno zmiennymi zależnymi. W tej pracy stosujemy technikę splotów, a rozkłady dochodów przybliżamy trójparametrycznym modelem Daguma.

Słowa kluczowe: rozkład dochodów, splot rozkładów, model Daguma

WSTĘP

Z badań Budżetów Gospodarstw Domowych z lat 2000-2010 wynika, że w grupie polskich gospodarstw domowych rodziny pozostające na utrzymaniu dwóch osób (w tym większość małżeństw) stanowią ok. 50%-52% wszystkich rodzin, a gospodarstwa będące na utrzymaniu jednej osoby ok. 28%-30%. Łącznie te dwie grupy stanowią przeważającą, ok. 80% grupę wszystkich rodzin. Interesujące jest zbadanie związku pomiędzy rozkładami dochodów tych dwóch grup gospodarstw domowych. Metoda badawcza przyjęta w niniejszej pracy wynika z prostej obserwacji. Dochód z gospodarstw domowych pozostających na utrzymaniu dwóch osób jest sumą dochodów indywidualnych x i y . Przyjmijmy, że

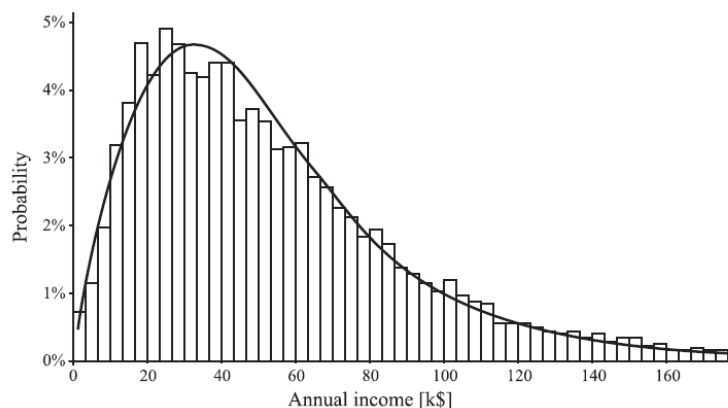
Z jest zmienną losową o wartościach z , a X i Y zmiennymi losowymi odpowiednio o wartościach x i y . Wówczas $Z = X + Y$. Przyjmijmy dalej, że f_X i f_Y są funkcjami gęstości zmiennych losowych X , Y i że zmienne te są niezależne. Wówczas, funkcja gęstości f_Z zmiennej losowej Z jest splotem funkcji f_X i f_Y oraz wyraża się wzorem [Plucińska, Pluciński 2000]

$$f_Z(z) = (f_X * f_Y)(z) = \int_R f_X(t) f_Y(z-t) dt. \quad (1)$$

Zauważmy, że równość $f_Z = f_X * f_Y$ nie implikuje niezależności zmiennych losowych, natomiast brak równości implikuje zależność tych zmiennych.

W przypadku dochodów amerykańskich obserwujemy, że rozkład dochodów rodzin z dwoma osobami pracującymi jest splotem rozkładów dochodów indywidualnych. Pokazano to w pracy [Drăgulescu, Yakovenko 2001] oraz potwierdzono w [Łukasiewicz, Orłowski 2004] dla danych dochodowych z roku 2000 – Rys. 1. Związek ten obserwuje się także w kolejnych latach, co wskazuje, że można go uważać za stałą cechę dochodów amerykańskich. Jednocześnie w pracy [Łukasiewicz, Orłowski 2004] pokazano, że w Polsce taka zależność nie istnieje. Model (1) jest wyraźnie przesunięty względem rozkładu empirycznego w kierunku wyższych dochodów – Rys. 2.

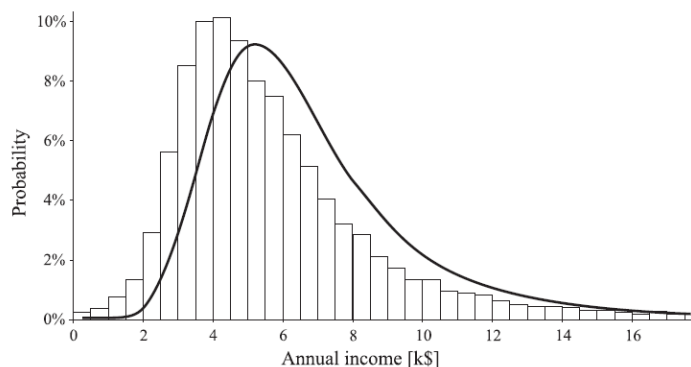
Rysunek 1. Empiryczny rozkład dochodów rodzin z dwiema osobami pracującymi w USA w roku 2000 oraz splot rozkładów indywidualnych (linia ciągła)



Źródło: [Łukasiewicz, Orłowski 2004]

W niniejszej pracy przedstawiono wyniki badań dla polskich gospodarstw domowych dla lat 2000–2010. Przeprowadzono analizy w dwóch zakresach: 1) pełnym, w odniesieniu do wszystkich rodzin utrzymywanych przez dwie osoby, 2) wąskim, dotyczącym tylko małżeństw bez dzieci.

Rysunek 2. Empiryczny rozkład dochodów rodzin z dwiema osobami pracującymi w Polsce w roku 2000 oraz splot rozkładów indywidualnych (linia ciągła)



Źródło: [Łukasiewicz, Orłowski 2004], przyjęto 1 USD = 4,143 zł

DANE I ORGANIZACJA BADANIA

Analizy oparto na danych dochodowych pochodzących z badań Budżetów Gospodarstw Domowych (BGD) z okresu 2000–2010. Zmiany w rozkładach dochodów obserwowane w kolejnych latach są niewielkie. Dane pobrano z odstępem dwuletnim, biorąc pod uwagę sześć lat parzystych. Dostępne dane dochodowe to miesięczne dochody rozporządzalne gospodarstw domowych, tzn. dochody netto, które przeznaczone są na wydatki oraz przyrost oszczędności. Dochód rozporządzalny każdego gospodarstwa jest równy sumie dochodów netto wszystkich jego członków. Dane pozwalają na analizowanie łącznego dochodu gospodarstwa (ewentualnie przeliczonego na osobę); w bazach danych BGD dochody poszczególnych członków gospodarstwa nie są udostępniane. Przed analizą dochody miesięczne przeliczone zostały na dochody roczne.

Gospodarstwa podzielono na grupy przyjmując za jedyne kryterium liczbę osób wnoszących dochód do gospodarstwa. Osoby te to nie tylko osoby pracujące, ale również osoby posiadające dochody w formie emerytury, renty, zasiłków, dochodów z wynajmu lokali, itp. W prezentowanym badaniu rozpatrywano następujące zbiory gospodarstw domowych:

- $R^{(2)}$ – rodziny utrzymywane przez dwie osoby;
- M – małżeństwa bez dzieci, w których każdy z małżonków osiąga dochody;
- $R^{(1)}$ – rodziny utrzymywane przez jedną osobę;
- S – gospodarstwa jednoosobowe.

Oczywiście $M \subset R^{(2)}$ oraz $S \subset R^{(1)}$.

Zbiory $R^{(1)}$ i S posłużyły konstrukcji rozkładów dochodów indywidualnych. Ponieważ rozkłady dochodów kobiet i mężczyzn wykazują we wszystkich latach duże różnice (Rys. 3) dokonano, więc dodatkowych podziałów wg płci:

- $R^{(1)} = R_k^{(1)} \cup R_m^{(1)}$, gdzie $R_k^{(1)}$, $R_m^{(1)}$ – rodziny utrzymywane odpowiednio przez kobiety, mężczyznę;
- $S = S_k \cup S_m$, gdzie S_k , S_m – samotne kobiety, samotni mężczyźni.

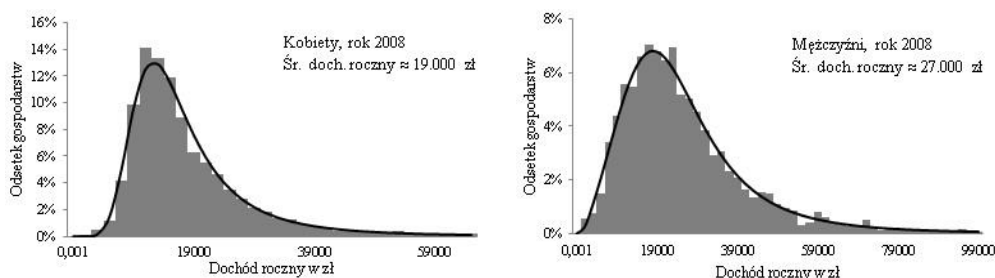
Liczebności wszystkich prób zamieszczono w Tabeli 1.

Tabela 1. Liczebności prób

Rok	$R_k^{(1)}$	$R_m^{(1)}$	$R^{(2)}$	S_k	S_m	M
2000	5.638	4.661	18.357	3.905	1.185	5.854
2002	5.215	4.438	15.460	3.671	1.377	5.080
2004	5.277	4.466	15.578	3.654	1.336	5.008
2006	6.102	4.840	18.095	4.377	1.644	6.248
2008	6.176	4.468	18.111	4.775	1.774	6.938
2010	6.106	4.804	18.472	4.804	1.856	7.507

Źródło: obliczenia własne

Rysunek 3. Rozkłady dochodów kobiet i mężczyzn w roku 2008 (rozkłady dochodów gospodarstw utrzymywanych przez kobiety $R_k^{(1)}$ i mężczyznę $R_m^{(1)}$). Linia ciągła – wykres funkcji aproksymującej (2)



Źródło: opracowanie własne

Przebieg badania dla każdego roku był następujący.

- 1) W oparciu o zbiory $R_k^{(1)}$, $R_m^{(1)}$ skonstruowano unormowane rozkłady dochodów kobiet i mężczyzn. Rozkłady te aproksymowano następnie funkcją gęstości Daguma, która określona jest wzorem

$$f(x) = \frac{abc}{x^{b+1}(1+ax^{-b})^{c+1}}, \quad (2)$$

gdzie $a > 1$, $b > 0$, $c > 0$ są parametrami modelu, a $x > 0$ jest dochodem.

Model (2) opisany jest w pracach [Dagum 1977] i [Łukasiewicz, Orłowski 2003]. Charakteryzuje się bardzo wysoką zgodnością z empirycznymi rozkładami dochodów. Stosowany jest często w empirycznych badaniach dochodów, por. [Dagum, Lemmi 1987], [Stoppa 1995], [Bandourian et al 2002], [Quintano, D'Agostino 2006]. Parametry funkcji (2) oszacowano metodą największej wiarygodności w oparciu o dane indywidualne [por.

Łukasiewicz, Orłowski 2003]. Otrzymano w ten sposób teoretyczne rozkłady dochodów indywidualnych kobiet i mężczyzn: f_k oraz f_m . Stopień zgodności z rozkładami empirycznymi zmierzono za pomocą wskaźników SSE i W_r [Bandourian et al 2002]:

$$SSE = \sum_{i=1}^k \left[\frac{n_i}{n} - p_i \right]^2, \quad (3)$$

$$W_r = \left(1 - \frac{1}{2} SAE \right) \times 100\%, \quad SAE = \sum_{i=1}^k \left| \frac{n_i}{n} - p_i \right|, \quad (4)$$

gdzie n_i / n oznacza częstości empiryczne, p_i częstości teoretyczne, natomiast k jest liczbą przedziałów histogramu.

- 2) Wyznaczono spłot rozkładów f_k i f_m zgodnie z formułą (1), który w przypadku funkcji Daguma i przy założeniu $x > 0$ przybiera postać

$$(f_k * f_m)(x) = \int_0^x \frac{a_1 b_1 c_1 \cdot a_2 b_2 c_2}{t^{b_1+1} (x-t)^{b_2+1} (1+a_1 t^{-b_1})^{c_1+1} (1+a_2 (x-t)^{-b_2})^{c_2+1}} dt, \quad (5)$$

gdzie a_1, b_1, c_1 oraz a_2, b_2, c_2 to parametry oszacowanych modeli, a $x > 0$ jest dochodem. Całkę (5) obliczano numerycznie.

- 3) Model $f_k * f_m$ porównano z empirycznym rozkładem dochodów rodzin grupy $R^{(2)}$. Wykorzystano w tym celu wskaźniki SSE i W_r .
- 4) Powtórzono 1) i 2) dla grup S_k, S_m , a otrzymany spłot porównano z rozkładem dochodów małżeństw bez dzieci.

WYNIKI

Rozkłady dochodów indywidualnych przybliżono za pomocą modeli Daguma (2). Łącznie oszacowano 24 funkcje, po 4 dla każdego roku. Nie zamieszczamy szczegółowych informacji o poszczególnych modelach, nadmienimy jedynie, że w prezentowanym badaniu wartości wskaźnika SSE obserwowano w granicach 0,00042 – 0,0013, a wskaźnika W_r w granicach od 93% do 96%. Oszacowane modele charakteryzują się wysoką i bardzo wysoką zgodnością z rozkładami empirycznymi.

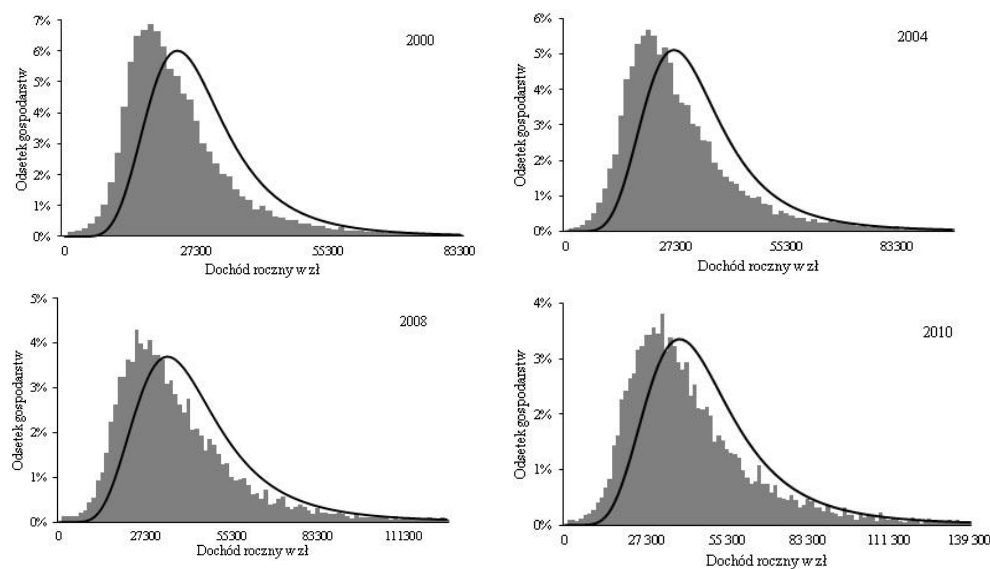
Wykorzystując formułę (5) dokonano spłotu funkcji Daguma przybliżających rozkłady dochodów kobiet i mężczyzn (grupy $R_k^{(1)}, R_m^{(1)}$). Otrzymano w ten sposób modele dochodów rodzin będących na utrzymaniu dwóch osób. Wyniki porównano z rozkładami empirycznymi (grupa $R^{(2)}$). Rezultaty dla wybranych lat pokazano na Rys. 4. W Tabeli 2 zamieszczono wartości

wskaźników zgodności SSE i W_r , których zmiany w czasie pokazano również na Rys. 6.

We wszystkich latach mamy do czynienia ze znacznym przesunięciem rozkładów teoretycznych w kierunku wyższych dochodów, natomiast kształt wykresów w przybliżeniu odpowiada rozkładowi empirycznemu. Ogólnie można powiedzieć, że dochody teoretyczne rodzin będących na utrzymaniu dwóch osób są wyższe od empirycznych. Jednocześnie, w całym okresie 2000-2010 obserwujemy spadek wskaźnika SSE i wzrost W_r , co wskazuje na systematyczne zmniejszanie się różnicy pomiędzy splotem, a rozkładem empirycznym.

W drugim etapie badań dokonano zawężenia danych, wykluczając dodatkowe, nieprzynoszące dochodu osoby w gospodarstwach. Dochody rodzin z jedną osobą pracującą zastąpiono dochodami rodzin jednoosobowych kobiet i mężczyzn (grupy S_k , S_m). Splot funkcji Daguma tych rozkładów porównano z rozkładem dochodów małżeństw bez dzieci. Wyniki przedstawiono na Rys. 5, w Tabeli 2 i na Rys. 7. W przypadku dochodów małżeństw również obserwujemy przesunięcie rozkładu teoretycznego w kierunku wyższych dochodów, ale znacznie mniejsze niż w przypadku ogólnym. Wartości wskaźnika SSE są znacznie niższe, a wartości wskaźnika W_r osiągają granicę 90%. Splot rozkładów indywidualnych nie opisuje dokładnie rozkładu dochodów małżeństw, jednak można uznać, że jest do niego bardzo zbliżony. Jednocześnie widzimy, że niezgodność ta maleje w czasie.

Rysunek 4. Rozkłady dochodów rodzin na utrzymaniu dwóch osób oraz splot rozkładów dochodów indywidualnych – model (5)



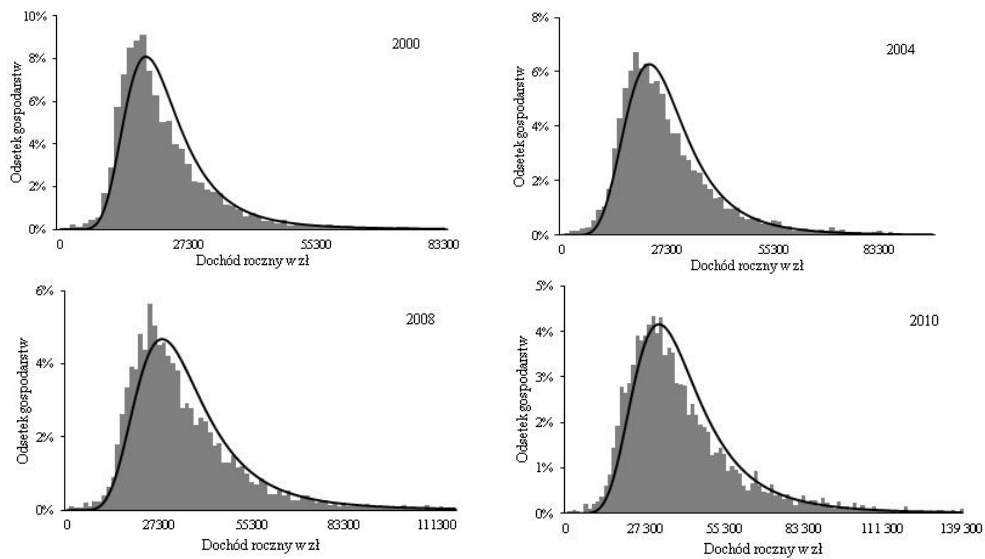
Źródło: opracowanie własne

Tabela 2. Zmiany wskaźników SSE i W_r

Rok	Rodziny na utrzym. dwóch osób		Małżeństwa bez dzieci	
	SSE	W_r	SSE	W_r
2000	0,0089	76,1%	0,0028	88,2%
2002	0,0070	77,9%	0,0020	89,0%
2004	0,0063	78,2%	0,0013	90,4%
2006	0,0058	77,8%	0,0021	88,4%
2008	0,0036	80,8%	0,0015	89,4%
2010	0,0031	81,5%	0,0010	90,1%

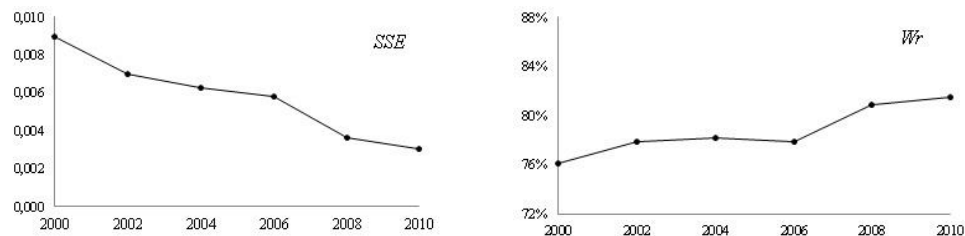
Źródło: obliczenia własne

Rysunek 5. Rozkłady dochodów małżeństw bez dzieci oraz splot rozkładów dochodów indywidualnych – model (5)



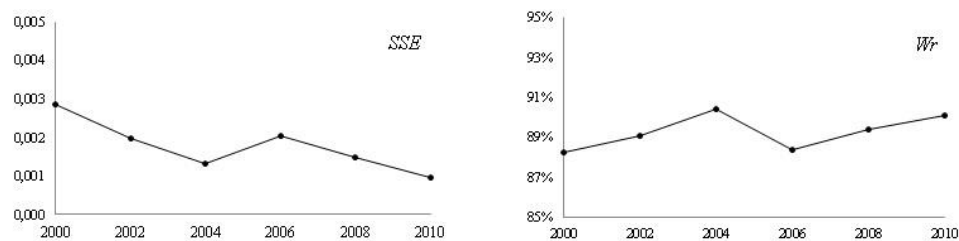
Źródło: opracowanie własne

Rysunek 6. Zmiany wskaźników SSE i W_r . Stopień zgodności rozkładów dochodów rodzin na utrzymaniu dwóch osób z modelem (5)



Źródło: opracowanie własne

Rysunek 7. Zmiany wskaźników SSE i W_r . Stopień zgodności rozkładów dochodów małżeństw bez dzieci z modelem (5)



Źródło: opracowanie własne

PODSUMOWANIE

W pracy porównano dochody gospodarstw domowych utrzymywanych przez dwie osoby ze splotem dochodów gospodarstw z jedną osobą zarabiającą. Zaobserwowano wyraźne, lecz malejące w okresie 2000 – 2010 różnice. Dochody otrzymane w wyniku splotu są wyższe dla wszystkich lat. Posiadanie rodziny skutkuje zmniejszeniem się łącznych dochodów, prowadzi do pewnej degradacji dochodowej. Otrzymane wyniki jednoznacznie wskazują, iż dochody członków rodzin w Polsce są zależne. Pokazano, że stopień tej zależności jednak systematycznie maleje.

Poszukując przyczyny zaobserwowanych różnic dokonano zawężenia danych, wykluczając nieprzynoszące dochodu osoby w gospodarstwach. Dochody małżeństw bez dzieci porównano ze splotem dochodów gospodarstw jednoosobowych. Również w tym przypadku dochody otrzymane dla splotu są wyższe w całym analizowanym okresie czasu. Zaobserwowane różnice są jednak wyraźnie mniejsze. Dodatkowo obserwujemy zmniejszanie się tych różnic na przestrzeni lat.

Otrzymane wyniki wskazują, że zawarcie małżeństwa i założenie wspólnego gospodarstwa domowego skutkuje pewnym zmniejszeniem łącznych dochodów. Ich dalsza degradacja następuje, gdy wzrasta liczebność gospodarstwa domowego.

Z drugiej strony, zaobserwowane różnice dochodowe na przestrzeni lat stopniowo maleją.

BIBLIOGRAFIA

- Bandourian R., McDonald J.B., Turley R.S. (2002) A comparison of parametric models of income distribution across countries and over time, Department of Economics, Brigham Young University.
- Dagum C. (1977) A new model of personal income distribution: Specification and Estimation. *Economic Appliquee* XXX (3), 413-437.
- Dagum C., Lemmi A. (1987) A Contribution to the Analysis of Income Distribution and Income Inequality and a Case Study: Italy, *Econometric Society Meeting*, Chicago.
- Drăgulescu A., Yakovenko V.M. (2001) Evidence for the exponential distribution of income in the USA. *The European Physical Journal B* 20, 585-589.
- Łukasiewicz P., Orłowski A. (2004) Probabilistic models of income distributions, *Physica A* 344, 146-151.
- Łukasiewicz P., Orłowski A. (2003) Probabilistic models of income distributions of polish households, *Metody ilościowe w badaniach ekonomicznych III*, Wyd. SGGW, 122-130.
- Plucińska A., Pluciński E. (2000) *Probabilistyka*, WNT, Warszawa.
- Quintano C., D'Agostino A. (2006) Studying inequality in income distribution of single-person households in four developed countries. *Review of Income and Wealth*, 52, 525-546.
- Stoppa G. (1995) Explicit estimators for income distributions, in: Dagum C., Lemmi, A. (eds.): *Research on Economic Inequality, Vol. 6: Income Distribution, Social Welfare, Inequality and Poverty*. Greenwich, CT: JAI Press, 393-405.

PERSONAL INCOMES VS HOUSEHOLDS INCOMES IN POLAND IN THE YEARS 2000 – 2010

Abstract: In this paper we study dependencies between distributions of personal and households incomes in Poland in 2000 to 2010. For example, the distributions of households incomes in the USA can be obtained by a convolution of distributions of personal incomes of family members. As was derived from our previous studies for 2004, there is no such a simple dependency in Poland. In principle, there is a very little knowledge about a type of relationship between those incomes in Poland. In the case of the USA personal incomes of family members are probably independent of each other. On the other hand, those variables are for sure dependent for Poland. In this paper we study personal and households incomes in Poland, their changes in time and their mutual relationship. Incomes distributions have been described by the three-parameter Dagum function.

Keywords: income distribution, convolution of distributions, Dagum's model

EKSPORT A DYNAMIKA PRODUKCJI SEKTORÓW ROLNYCH KRAJÓW UE

Jacek Strojny

Katedra Statystyki Matematycznej, Uniwersytet Rolniczy w Krakowie
e-mail: rrstrojn@cyf-kr.edu.pl

Streszczenie: Opracowanie ma na celu śledzenie dynamicznego sprzężenia między eksportem a produkcją sektora agrobiznesu w krajach UE. Dokonano tego poprzez badanie oszacowań korelacji między wskazanymi zmiennymi oraz przez obserwację dynamiki wskaźnika relacji produkcja / eksport sektorów rolnych krajów UE w latach 1994–2010. Rezultaty studium wskazują na występowanie zależności między badanymi zmiennymi oraz ukazują szeroki zakres zróżnicowania wzorców relacji między eksportem rolnym a produkcją końcową rolnictwa w badanych krajach.

Słowa kluczowe: eksport rolny, produkcja rolna, kraje UE

WPROWADZENIE

Rosnący od kilku lat polski eksport na rynki krajów UE postrzegany jest jako czynnik dynamizacji rozwoju rodzimego sektora rolno-żywnościowego. Poszerzenie rynku w wyniku zniesienia ograniczeń eksportowych przyciągnęło do sektora agrobiznesu znaczące inwestycje, co przyczyniło się do wyraźniej poprawy konkurencyjności polskich produktów żywnościowych na arenie międzynarodowej. Zwiększanie wolumenu sprzedaży żywności i pozostałych produktów pochodzenia rolniczego za granicę otwiera przed krajowymi producentami szanse rozszerzania skali produkcji [Xiao, Reed 2007]. Helpman i Krugman [1985] argumentują, że wzrost w gospodarce implikuje rozwój eksportu opartego na malejących kosztach wraz ze wzrostem skali produkcji. Autorzy wskazują również na eksport jako sposób obejścia bariery rynku wewnętrznego.

Wzrost konsumpcji, rozwój rynku stymuluje zwiększenie produkcji rolnej. Jednak w krajach wysokorozwiniętych produkty żywnościowe w przewadze charakteryzują się niską elastycznością dochodową popytu, co znacznie ogranicza perspektywę zwiększania produkcji wynikającej z rosnącego dochodu. Stabilizacja

rynków krajowych nie przekreśla jednakże możliwości rozszerzania produkcji rolnej w wyniku ekspansji na rynek międzynarodowy. Z jednej strony perspektywy zbytu produktów na rynkach zagranicznych tworzą warunki do wzrostu produkcji krajowej, ale równocześnie czynią ją bardziej wrażliwą na wahania popytowo-podażowe rynku światowego. Z tego tytułu skutki kryzysu finansowego zapoczątkowanego w 2008 r. znalazły odzwierciedlenie (nie tak silne, jak w przypadku towarów przemysłowych) w dynamice rynkowej większości grup asortymentowych branży spożywczej. Pogorszenie się ogólnego klimatu ekonomicznego w gospodarce światowej rzutowało na poziomy konsumpcji, rozmiary handlu produktami żywnościowymi oraz ich ceny. Szczególnie dotkliwe skutki wywołały spadki cen, których nawet nieznaczne wahania wpływają na sytuację dochodową producentów. Problem ten ma większe znaczenie dla sektora rolnego w UE niż dla innych znaczących eksporterów na rynek światowy ze względu na strukturę i poziom kosztów produkcji. Niejednokrotnie, łączne subwencje i wsparcie polityki publicznej osiąga w poszczególnych krajach Unii wartości wyższe od dochodu netto gospodarstw rolnych.

W związku z dynamicznym rozwojem eksportu żywności dominują powszechne oczekiwania dalszego wzrostu obrotów polskiego handlu zagranicznego produktami rolnymi i podtrzymania wynikającego z tego tytułu stymulacyjnego wpływu efektu popytowego na rozwój produkcji krajowej. Jednak nie należy ignorować zasad i specyfiki warunków produkcji rolnej w Unii Europejskiej – przede wszystkim zakresu regulacji i wsparcia przez politykę publiczną rolnictwa. Morley i Morgan [2008] dowiedli, że nierówne poziomy wsparcia stosowane przez poszczególne kraje UE doprowadziły do strukturalnych nadwyżek (nie wynikających z produktywności) w obrocie produktami rolnymi w niektórych krajach podczas gdy inne odnotowują długofalowe deficyty.

Mimo wdrożenia zaostrzonego systemu kontroli jakości artykułów rolno-spożywczych w Polsce (mającego na celu także zwiększenie konkurencyjności rodzimych produktów) [Cymanow 2009], ekspansję polskiej żywności na rynki Unijne mogą wydatnie ograniczyć skandale związane z zanieczyszczeniem żywności, które podkopują zaufanie nabywców do produktów importowanych oraz mogą wywoływać działania prewencyjne władz poszczególnych krajów członkowskich mające na celu ochronę konsumentów. Także stopniowe zanikanie przewag konkurencyjnych (głównie kosztowych) polskich producentów wiąże się z koniecznością poszukiwania nowych rynków zbytu (poza Unią Europejską), jako warunku podtrzymania dynamiki gospodarki rolno-żywnościowej [Herzer, Nowak–Lehmann 2006].

Celem opracowania jest śledzenie dynamicznej współzależności między eksportem a produkcją sektora rolnego w krajach UE. Studium ma charakter wstępnej eksploracji charakteru związków i zakresu ich zróżnicowania wśród rozpatrywanych krajów. Modelowaniu poddano zmienne: wielkość eksportu sektora rolnego oraz wartość produkcji końcowej rolnictwa.

EKSPORT A POTENCJAŁ WZROSTOWY GOSPODARKI

Branża nakierowana na eksport ma potencjał dostarczenia dla gospodarki w wielokierunkowy sposób impulsu wzrostowego. Poprzez łańcuch powiązań między tym sektorem a gałęziami z nim kooperującymi kreowane są możliwości rozwoju „ciągnionego”. Szybko rozwijająca się gałąź, która celuje w lokowaniu swojej produkcji na rynku międzynarodowym dodatkowo kreuje w gospodarce przyjazne środowisko do wzmożonego inwestowania. Dzięki przyjaznemu klimatowi tworzonymu przez intensywnie rozwijający się sektor nakierowany na eksport również przedsiębiorstwa z innych gałęzi polepszają swoje szanse rozwoju korzystając z dobrej koniunktury (z uwagi na komplementarny charakter powiązań w gospodarce).

Programy wzrostu gospodarczego bazujące na eksporcie mają w historii gospodarki światowej długie tradycje – już w XIX w. taka polityka ekonomiczna była wdrażana przez kraje Ameryki Łacińskiej. Model wzrostu gospodarczego czerpiący z intensywnego rozwoju wybranych sektorów gospodarki, które swą dynamikę zawdzięczają eksportowi był realizowany z sukcesem po II wojnie światowej przez wiele społeczeństw. Taka koncepcja rozwoju zakłada intensywne zwiększanie eksportu, w tym eksportu per capita, powiązane ze wzrostem produktywności pracy w sektorze nakierowanym na eksport.

Jednak, warunkiem przeniesienia impulsów wzrostowych na resztę gospodarki jest zaistnienie efektu „rozlania” poprawy produktywności w sektorze eksportującym na inne gałęzie. Skutki rozprzestrzeniania się efektów stymulacyjnych sektora nakierowanego na eksport na pozostałe działy gospodarki zależą w dużej mierze od struktury danej gospodarki, a w związku z tym od charakteru powiązań międzygałęziowych. Poprawa średniej produktywności wszystkich działów – nie tylko w sektorze proeksportowym – zwykle stanowi impuls przyciągający do danej gospodarki narodowej nowe inwestycje. Spełnienie łącznie powyższych warunków pozwala na upatrywanie w rozwoju sektora eksportującego źródła wzrostu całości gospodarki narodowej. Zwykle, efektem procesów wzrostowych zachodzących w gospodarce we wskazanej sekwencji jest również realny przyrost dochodu per capita.

Struktura produkcji danego kraju w znacznym zakresie jest determinowana zasobami naturalnymi, poziomem technologii i ogólnym stopniem rozwoju społeczno-gospodarczego, a także wielkością rynku krajowego. Ostatnie z wymienionych ograniczeń wzrostu – barierę pojemności rynku krajowego można pokonać poprzez ekspansję eksportową. Charakter bodźców wzrostowych wywodzonych z sektora nakierowanego na eksport jest przejawem schematu współzależności, które występują między działami w danej gospodarce. Zatem, uprawnione wydaje się stwierdzenie, że zarówno polityka handlowa, jak i strategie eksportowe są determinowane przez strukturę gospodarczą poszczególnych krajów.

METODA I DANE

Mimo powszechnego przekonania o pozytywnym wpływie aktywności eksportowej na rozwój produkcji krajowej, na proces wzrostu gospodarczego [Balassa 1978] ujęcie tych zależności w formie modelowej nasyca wiele trudności, a studia empiryczne dostarczają wieloznaczne wyniki. Źródłem dodatkowych zakłóceń (w relacji do mniej regulowanych działów gospodarki) w badaniach procesów rozwojowych rolnictwa jest zakres ingerencji polityki publicznej. Carlin i inni [1997] w oparciu o badania krajów OECD oceniają, że jedynie w długim okresie (kilkunastu lat) można poszukiwać współzależności między zmianami efektywności w produkcji oraz ich odzwierciedleniem w postaci ewolucji udziałów w rynku międzynarodowym. Jednak tych spostrzeżeń nie potwierdzają Meeusen i Rayp [2000]. Gurgul i Lach [2010] w poszukiwaniu czynników wzrostu polskiej gospodarki dokonali przeglądu liniowych i nieliniowych metod wykrywania zależności między handlem zagranicznym a wzrostem gospodarczym. W badaniu obejmujących lata 1996-2009 autorzy uzyskali pozytywny wynik testów przyczynowości dla związków o charakterze liniowym między tempem wzrostu PKB i tempem wzrostu eksportu niezależnie od rozpatrywanego podokresu badania. Natomiast analiza metodami nieliniowymi pozwoliła jedynie na stwierdzenie o słabej zależności między badanymi procesami.

Zarówno ujęcia teoretyczne hipotezy o pozytywnym wpływie eksportu na wzrost gospodarczy, jak i próby weryfikacji empirycznej tego stwierdzenia wprowadzają wiele niejednoznaczności. Wyniki studiów empirycznych bywają źródłem sprzecznych wniosków. Ogólnie, ujęcia metodologiczne problemu weryfikacji hipotezy o pozytywnym wpływie eksportu na wzrost gospodarczy bądź zwiększanie bogactwa społecznego poprzez rozszerzanie skali produkcji wybranych sektorów gospodarki oparte o dane przekrojowo-czasowe można sklasyfikować w trzy nurty:

- testy hipotezy o stymulacyjnej roli eksportu w procesie wzrostu gospodarczego z użyciem miar korelacji,
- modele oparte na regresji szacowanej metodą najmniejszych kwadratów,
- koncepcje wykorzystujące różnorakie techniki analizy szeregów czasowych.

W niniejszym opracowaniu punktem wyjścia do oceny wpływu eksportu na wielkość produkcji rolnej były miary korelacji między rozważanymi zmiennymi. Zbiór danych wykorzystany w badaniu obejmował informacje o rozmiarach eksportu i importu towarów opartych na surowcach rolnych oraz wartość produkcji końcowej rolnictwa krajów członkowskich UE za lata 1994-2010. Na podstawie danych źródłowych oszacowano wartość eksportu netto produktów rolnych jako różnicę między eksportem i importem. Dla każdego z obiektów-krajów i okresów wyznaczono również wskaźnik relacji produkcja/eksport, który posłużył do oceny tempa ewolucji wzmiankowanej zależności. Szacunku tego dokonano

z wykorzystaniem trendu liniowego gdzie za zmienną zależną przyjęto wartość wskaźnika, a jego dynamikę rozpatrywano jako funkcję czasu.

Celem zweryfikowania i oszacowania w badanym zbiorze państw efektu przyrostów produkcji rolniczej wpływających z eksportu oszacowano parametry modelu regresji dla danych przekrojowo-czasowych. Model z efektami ustalonymi przyjął postać:

$$y_{it} = \beta x_{it} + u_{it}, \quad \text{dla } i = 1 \dots, N, \quad t = 1 \dots, T, \quad (1)$$

gdzie dla obiektów i , okresów t :

y_{it} – obserwacja zmiennej zależnej – produkcja końcowa rolnictwa,

x_{it} – zmienna niezależna obiektu i w okresie t – eksport produktów sektora agrobiznesu,

β_{it} – $k \times 1$ wektor parametrów zmiennych niezależnych,

u_{it} – składnik błędu obiektu i w okresie t .

Model z efektami ustalonymi zakłada dekompozycję składnika błędu u_{it} :

$$u_{it} = \alpha_i + \varepsilon_{it}, \quad (2)$$

gdzie:

α_i – stały w czasie efekt grupowy,

ε_{it} – składnik błędu właściwy obserwacji.

Składnik α_i jest następnie traktowany jako stały parametr, właściwy dla obiektu (stała w modelu regresji). Zatem:

$$y_{it} = \beta x_{it} + \alpha_i + \varepsilon_{it}, \quad \text{dla } i = 1 \dots, N, \quad t = 1 \dots, T. \quad (3)$$

Zakładamy, że $\alpha_i \sim N(0, \delta_\alpha^2)$, $\varepsilon_i \sim N(0, \delta_\varepsilon^2)$ oraz α_i i ε_i są niezależne.

EKSPORT A PRODUKCJA ROLNICZA KRAJÓW UE

Kryzys gospodarczy, efekty załamania się systemu finansowego z 2008 roku w rolnictwie są odczuwane w mniejszym stopniu w postaci dramatycznych zmian popytu, jednak obserwowane są wyraźne obniżki cen większości płodów rolnych. Dodatkowo, wzrost cen podstawowych środków produkcji, maszyn powoduje pogorszenie się opłacalności działalności gospodarczej sektora agrobiznesu – przede wszystkim rolnictwa. Ograniczenia kredytowe dla przedsiębiorstw przetwórczych przemysłu rolno-spożywczego mogą być źródłem dodatkowych bodźców hamujących produkcję sektora. W następstwie niekorzystnej sytuacji na rynkach produktów rolnych zmniejszyła się skłonność do inwestowania w sektorze agrobiznesu [Musiał 2009]. Mimo znamion kryzysu gospodarczego w państwach będących głównymi rynkami dla polskich towarów rolno-spożywczych eksport tych artykułów wykazuje znaczną dynamikę wzrostu. Ogólne pogorszenie się sytuacji ekonomicznej w gospodarce światowej znalazło jednak odzwierciedlenie w poziomach, zarówno produkcji, jak i wymiany międzynarodowej produktami rolno-żywnościowymi.

Tabela 1. Charakterystyki relacji między eksportem sektora a produkcją końcową rolnictwa krajów UE w latach 1994-2013

L.p.	Grupa	Kraj	Współczynnik korelacji Pearsona		Wskaźnik relacji produkcja / eksport		Oszacowania parametrów trendu liniowego dla wskaźnika relacji produkcja / eksport	
			eksport - produkcja	eksport netto - produkcja	rok 1994	rok 2010	β	stała
1	1	Austria	0,597	-0,029	4,13	0,65	-0,158	2,759
2	2	Belgia	0,945	0,823	0,34	0,25	-0,011	0,435
3	1	Bułgaria	0,884	0,213	1,74	1,12	-0,080	3,046
4	3	Cypr	-0,512	-0,881	1,23	2,85	0,176	0,403
5	1	Czechy	0,837	-0,713	3,15	0,88	-0,173	3,715
6	2	Dania	0,787	0,805	0,78	0,52	-0,019	0,811
7	1	Estonia	0,933	-0,569	1,55	0,80	-0,040	1,433
8	4	Finlandia	0,589	-0,414	3,68	1,35	-0,063	2,522
9	1	Francja	0,942	0,751	1,49	1,12	-0,026	1,515
10	3	Grecja	0,887	-0,873	4,63	3,16	-0,035	4,587
11	4	Hiszpania	0,949	0,718	2,47	1,27	-0,070	2,443
12	2	Holandia	0,778	0,800	0,45	0,24	-0,016	0,492
13	2	Irlandia	0,841	0,342	0,75	0,62	-0,014	0,790
14	1	Litwa	0,908	0,663	1,57	0,60	-0,120	2,730
15	2	Luksemburg	0,748	-0,927	0,46	0,20	-0,015	0,507
16	5	Łotwa	0,884	0,061	8,29	0,73	-0,366	5,960
17	4	Malta	0,500	-0,810	6,33	1,62	-0,216	4,435
18	1	Niemcy	0,905	0,110	1,52	0,79	-0,053	1,647
19	5	Polska	0,906	0,766	5,43	1,34	-0,325	6,428
20	4	Portugalia	0,901	-0,932	4,36	1,63	-0,180	4,497
21	6	Rumunia	0,822	-0,804	24,69	4,25	-0,970	24,282
22	1	Słowacja	0,707	-0,726	4,18	0,69	-0,242	4,609
23	1	Słowenia	0,695	-0,733	2,28	0,56	-0,149	3,103
24	1	Szwecja	0,669	-0,754	2,92	0,95	-0,120	2,640
25	1	Węgry	0,897	0,776	1,96	0,90	-0,052	2,030
26	1	Wielka Brytania	0,560	-0,502	1,63	1,10	-0,035	1,545
27	4	Włochy	0,868	-0,309	3,21	1,56	-0,100	3,142

Źródło: obliczenia własne na podstawie FAOSTAT

Współzależności między produkcją końcową rolnictwa i eksportem produktów wytwarzanych z surowców rolnych przyjmują w poszczególnych krajach UE różną formę. Dla większości państw członkowskich między wskazanymi zmiennymi można zaobserwować relewantne powiązania, których zakres i kierunek próbowano identyfikować za pomocą współczynnika korelacji

Pearsona (Tabela 1). Dla przeważającej liczby państw miara korelacji wskazuje na występowanie silnych bądź bardzo silnych współzależności między rozważanymi zmiennymi. Jedynie dla Cypru wskaźnik przyjmuje umiarkowaną wartość ujemną (-0,512). W przewadze, wysokie wartości współczynnika korelacji sugerują pozytywne oddziaływanie eksportu dla stymulacji krajowej produkcji rolniczej.

Badane kraje wykazują znacznie większe zróżnicowanie ze względu na poziom wskaźnika relacji wartość produkcji do eksportu rolnego oraz intensywność przekształceń tego wskaźnika w okresie badania niż z uwagi na przytoczone oszacowania wskaźnika korelacji między wielkością eksportu i produkcji. W Tabeli 1 przedstawiono wskaźniki relacji produkcja/eksport zarówno dla pierwszego z rozpatrywanych okresów i dla 2010 r. Dla każdego z obiektów oprócz Cypru oszacowanie β przyjmuje wartość ujemną, co odpowiada malejącemu udziałowi produkcji w obrotach towarami wytwarzanymi z surowców rolnych. Wynika to z dynamicznych wzrostów, równocześnie eksportu i importu.

Bardziej skomplikowany charakter zależności między produkcją końcową rolnictwa a zakresem ekspansji międzynarodowej wyłania się po uwzględnieniu w rachunku korelacyjnym kategorii „eksport netto”. Postać relacji między produkcją a eksportem netto produktów rolnych wykazuje znacznie większe zróżnicowanie niż oddziaływanie na produkcję końcową sektora eksportu nie korygowanej wartością importu. Niektóre z państw cechujących się wysokimi wartościami współczynnika korelacji dla zależności między eksportem i produkcją rolniczą w przypadku relacji produkcja/eksport netto uzyskują niskie wartości ujemne. W takich okolicznościach źródeł dynamicznego wzrostu eksportu należy upatrywać w przetwórstwie importowanych półproduktów i surowców. Strategie ekspansji eksportowej wysoko przetworzonych produktów żywnościowych bazują w tym przypadku na szybko rosnącym imporcie surowców rolnych o niskim stopniu przetworzenia. W niektórych wysokorozwiniętych krajach UE sektor agrobiznesu coraz wyraźniej ewoluje w kierunku operowania w działach o wyższych marżach i zyskach. Koncepcja ta może być postrzegana również jako poszukiwanie sposobów budowania przewagi konkurencyjnej na rynku międzynarodowym.

Rozpatrując wartość wskaźnika relacji produkcja końcowa rolnictwa/eksport oraz kierunek i nasilenie zmian tej miary w czasie wśród krajów UE można poszukiwać w tym względzie pewnych analogii w bardziej homogenicznych podzbiorach państw. Grupowania krajów UE ze względu na kierunek ewolucji wzmiankowanego wskaźnika i tempo jego zmiany (szacowane współczynnikiem kierunkowym trendu liniowego) dokonano analizą skupień, metoda Warda, miara odległości – kwadrat odległości euklidesowej. Z uwagi na wskazane zmienne wyróżniono pięć homogenicznych skupisk obiektów-krajów. Najliczniejsza grupa 1-sza liczy dwanaście państw. Wyróżniono także dwie podgrupy pięcioelementowe (gr. 2-ga i 4-ta), dwie podgrupy dwu obiektowe (3-cia i 5-ta) oraz jedną tworzoną przez pojedyncze państwo. Przynależność poszczególnych krajów do

wymienionych skupisk w Tabeli 1 oznaczono zmienną grupą. Efekty procesu klasyfikacji przedstawiono w Tabeli 2. Tabela ta zawiera wartości rozważanych zmiennych uśrednione dla wyróżnionych grup taksonomicznych.

Tabela 2. Wartości uśrednione (w homogenicznych grupach krajów wyróżnionych metodą taksonomiczną) wskaźników opisujących zależności między eksportem, eksportem netto i produkcją końcową rolnictwa

Grupa	Korelacja		Wskaźnik relacji produkcja / eksport			Wskaźnik relacji produkcja / eksport netto		
	eksport - produkcja	eksport netto - produkcja	rok 1994	rok 2010	tempo zmian relacji	rok 1994	rok 2010	tempo zmian relacji
1	0,79	-0,13	2,34	0,85	-0,10	-1,55	-0,41	0,20
2	0,82	0,37	0,56	0,37	-0,01	1,06	1,18	0,00
3	0,19	-0,88	2,93	3,01	0,07	-17,45	-3,82	1,20
4	0,76	-0,35	4,01	1,49	-0,13	33,31	-1,23	-0,40
5	0,90	0,41	6,86	1,04	-0,35	-16,36	0,72	1,33
6	0,82	-0,80	24,69	4,25	-0,97	-33,93	-19,14	2,18
Ogółem	0,76	-0,12	3,53	1,18	-0,13	1,91	-1,13	0,28

Źródło: obliczenia własne na podstawie FAOSTAT

OCENA WPŁYWU EKSPORTU NA PRODUKCJĘ KOŃCOWĄ ROLNICTWA KRAJÓW UE

Próby oceny siły wpływu eksportu na produkcję końcową rolnictwa dokonano z wykorzystaniem modelu regresyjnego z efektami ustalonymi (Tabela 3) dla danych przekrojowo-czasowych. Ponieważ zarówno szereg y_{it} zmiennej zależnej (produkcja końcowa), jak i x_{it} (eksport produktów sektora agrobiznesu) badane testami pierwiastka jednostkowego Dickeya–Fullera (DF) oraz Kwiatkowskiego–Phillipsa–Schmidta–Shina (KPSS) nie wykazywały stacjonarności, dokonano przekształcenia danych poprzez wyznaczenie ich pierwszych przyrostów. Przeobrażone zmienne (Δy_{it} , Δx_{it}) spełniały już warunek stacjonarności. Na pierwszych różnicach zmiennych estymowano model regresyjny (dla danych panelowych z efektami ustalonymi) produkcji końcowej sektora rolnego względem eksportu wytwarzanych przez niego towarów.

Współczynnik determinacji dla tego modelu przyjmuje wartość ($R^2 = 0,501$), także oszacowane parametry są wysoce istotne. Różnicowanie szeregów czasowych zapewniło również usunięcie autokorelacji składnika losowego (statystyka Durбина–Watsona = 1,91).

Model wywodzący wzrost produkcji rolnej z eksportu sugeruje wielkość tego wpływu na 0,8195 w odniesieniu do przyrostu zmiennej Δy_{it} o jednostkę (Tabela 3 – wsp. β dla zmiennej Δx_{it}). Przy tym nie znaleziono uzasadnienia do uwzględnienia w równaniu regresyjnym kolejnych opóźnień zmiennej niezależnej

(Δx_{it}). Najsilniejszym z rozpatrywanych czynnikiem objaśniającym poziomy przyrostów produkcji jest składnik autoregresyjny szeregu czasowego (przyrosty eksportu). Wysoka wartość parametru β dla zaprezentowanego modelu oraz wysoka istotność statystyczna, sugerują silne związki eksportu z poziomami produkcji rolnej. Należy jednak zauważyć, że w przedstawionej formie eksport wyjaśnia nie więcej niż połowę zmienności produkcji wśród krajów UE.

Tabela 3. Wyniki estymacji parametrów modelu wpływu eksportu na produkcję końcową rolnictwa krajów UE w latach 1994 – 2010 (dane panelowe)

Składnik modelu	Współczynnik	Błąd standardowy	<i>t-Studenta</i>	<i>p</i>
stała	-246,562	87,6870	-2,8118	0,00517
Δx_{it}	0,819513	0,04228	19,3810	<0,00001

Źródło: obliczenia własne w programie Gretl

Bezpośrednia interpretacja siły tego oddziaływania pociąga jednak ryzyko przeszacowania efektu eksportu z uwagi na uwzględnienie w danych surowych (wielkość eksportu) szerokiego spektrum czynników makroekonomicznych. Należy także podkreślić znaczne zróżnicowanie charakteru relacji między produkcją a eksportem rolnym zidentyfikowane w poszczególnych krajach na etapie analizy korelacyjnej. Dodatkowo, zachodzi konieczność uwzględnienia licznych uwag autorów opracowań z zakresu modelowania danych o charakterze przekrojowo-czasowym odnośnie rozbieżności oszacowań współczynników modeli regresyjnych w zależności od zastosowanej metody estymacji parametrów.

PODSUMOWANIE

Analiza współzależności między eksportem towarów opartych na surowcach wytwarzanych przez rolnictwo a produkcją końcową sektora rolnego w krajach UE (za wyjątkiem Cypru) wykazała występowanie w tym zakresie znaczących, a dla wielu krajów nawet bardzo silnych związków. Wstępne spostrzeżenia wynikające z analizy korelacyjnej na temat badanych zależności znalazły potwierdzenie i zostały skwantyfikowane w modelu regresyjnym dla danych przekrojowo-czasowych.

W okresie badania jedynie w przypadku Cypru nie stwierdzono malejącej relacji wartości produkcji rolnej do eksportu. Wiąże się to z tendencją ewolucji sektora agrobiznesu w kierunku przetwórstwa surowców importowanych oraz eksportu produktów przetworzonych.

Poziom rozwoju, charakter i wielkość produkcji sektora agrobiznesu charakteryzują się znacznym zróżnicowaniem w badanym zbiorze państw, co znajduje odzwierciedlenie w zróżnicowaniu współzależności między eksportem oraz eksportem netto a wartością produkcji rolnictwa.

BIBLIOGRAFIA

- Balassa B. (1978) Exports and economic growth: further evidence, *Journal of Development Economics*, vol. 5(2), str. 181 – 189.
- Carlin W., Glyn A, van Reenen J. (1997) Quantifying a Dangerous Obsession Competitiveness and Export Performance in an OECD Panel Industries, CEPR, Discussion Paper nr 1628, London.
- Cymanow P. (2009) Wpływ akcesji Polski do UE na jakość handlową wybranych grup artykułów żywnościowych, *Problemy Rolnictwa Światowego*, tom 9(24), str. 27 – 34.
- Gurgul H., Lach Ł. (2010) International Trade and Economic Growth in the Polish Economy, *Operations Research and Decisions*, vol. 20(3–4), str. 5 – 29.
- Helpman E., Krugman P. (1985) *Market Structure and Foreign Trade*, Cambridge, MIT Press, Cambridge, MA.
- Herzer D., Nowak-Lehmann F.D. (2006) What does export diversification do for growth? An econometric analysis, *Applied Economics*, vol. 38, str. 1825 – 1838.
- Meeusen W., Rayp G., 2000, Patents and Trademarks as Indication of International Competitiveness, [W:] *Competitiveness and Value of intangible Assets*, P. Buigues, A. Jacquemin, J.F. Marchipont (red.), Edwards Elgar, Cheltenham.
- Morley B., Morgan W. (2008) Causality between Exports, Productivity and Financial Support in European Union Agriculture, *Regional Studies*, vol. 42(2), str. 189 – 198.
- Musiał W. (2009) Zagrożenie kryzysem i próby działań antykryzysowych w rolnictwie polskim, *Roczniki Naukowe SERiA*, vol. XI(2), str. 168 – 173.
- Xiao Q., Reed M. (2007) Export and production growth: evidence from three major wheat exporters of Australia, Canada and the United States, *Applied Economics*, vol. 39(1), str. 309 – 319.

**EXPORT AND DYNAMICS
OF EU COUNTRIES' AGRICULTURAL OUTPUT**

Abstract: The paper attempts to investigate the dynamic causality between exports and output of agribusiness sector of EU countries. This was achieved by examining correlation estimates concerning variables mentioned above and exploring production / exports ratio indicator dynamics in the period of 1994–2010. The study results showed a dependence between variables under investigation and revealed significant patterns differentiation of interdependence between agricultural exports and production in EU countries.

Keywords: agricultural export, agricultural production, EU countries

MODELE HARMONICZNE ZE ZŁOŻONĄ SEZONOWOŚCIĄ W PROGNOZOWANIU SZEREGÓW CZASOWYCH Z LUKAMI SYSTEMATYCZNYMI

Maria Szmuksta-Zawadzka, Jan Zawadzki

Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

e-mail: jan.zawadzki@zut.edu.pl

Streszczenie: W modelowaniu zmiennych ze złożoną sezonowością dla pełnych danych i danych z lukami niesystematycznymi mogą być wykorzystywane zarówno modele ze zmiennymi zero-jedynkowymi jak i modele harmoniczne. Natomiast w przypadku występowania luk systematycznych- jedynie oszczędne modele harmoniczne. W modelach tych każdy rodzaj wahań opisywany jest za pomocą odrębnych zestawów składowych sinuso- i kosinusoidalnych. Rozważania teoretyczne zostaną zilustrowane przykładem empirycznym.

Słowa kluczowe: modele harmoniczne, sezonowość złożona, brakujące dane

WPROWADZENIE TEORETYCZNE

Celem pracy jest wykazanie, że modele harmoniczne mogą być wykorzystywane w prognozowaniu zmiennych ze złożoną sezonowością w warunkach występowania systematycznych luk w danych. Złożoną sezonowość dla dziennych szeregów czasowych bez luk, uwzględniającą występowanie wahań o cyklu 12 miesięcznym oraz siedmiodniowym można opisać za pomocą odrębnych zespołów zmiennych zero-jedynkowych lub wielomianów trygonometrycznych (por. [Kufel 2010], [Szmuksta-Zawadzka, Zawadzki 2013]).

Model ze złożoną sezonowością z wielomianami trygonometrycznymi o cyklach 12-miesięcznym (rocznym) i 7-dniowym (tygodniowym) przyjmuje postać:

$$\begin{aligned}
 Y_t = & \alpha_1 t + \alpha_0 + \sum_{i=1}^6 (\alpha_{0i} M \sin \omega_i t_m + b_{0i} M \cos \omega_i t_m) \\
 & + \sum_{j=1}^3 (\alpha_{0j} D \sin \omega_j t + b_{0j} D \cos \omega_j t) + U_t
 \end{aligned} \tag{1}$$

gdzie:

$$\omega_i = \frac{2 \cdot \Pi \cdot i}{12} \quad (i = 1, 2, \dots, 6),$$

$$\omega_j = \frac{2 \cdot \Pi \cdot j}{7} \quad (j = 1, 2, 3),$$

$$t_m = 1, 2, \dots, 12, \dots, 12 \cdot r \quad (r - \text{liczba lat}),$$

$$t = 1, 2, \dots, n,$$

Symbole M i D poprzedzające zapisy funkcji trygonometrycznych odnoszą się odpowiednio do miesięcy i dni tygodnia.

W zapisie modelu (1) występują dwie zmienne oznaczające czas: t_m oraz t . Zmienna t_m przyjmuje takie same wartości dla wszystkich dni wchodzących w skład danego miesiąca. Takie jej zdefiniowanie oznacza, że wartości składowych harmonicznych w danym miesiącu będą identyczne. Tym samym więc model z wielomianem trygonometrycznym będzie równoważny modelowi ze zmiennymi zero-jedynkowymi. Równoważność ta polega na tym, że mając oceny parametrów jednego z nich można wyznaczyć parametry drugiego. Ponadto obydwa modele, dla pełnych danych, charakteryzują się identycznymi ocenami parametrów struktury stochastycznej oraz dają identyczne prognozy.

W szeregach czasowych z sezonowością niezłożoną mogą występować dwa rodzaje luk:

- niesystematyczne,
- systematyczne.

Występowanie luk niesystematycznych oznacza, że dostępna jest przynajmniej jedna obserwacja dla każdego podokresu.

Natomiast z lukami systematycznymi będziemy mieć do czynienia wtedy, gdy brak będzie danych przynajmniej o jednym podokresie.

W przypadku złożonej sezonowości możliwych kombinacji występowania luk jest więcej. Luki systematyczne lub niesystematyczne mogą występować bądź tylko dla jednego rodzaju wahań lub dla obu rodzajów wahań jednocześnie. Można też wyróżnić przypadek mieszany, w którym dla jednego rodzaju wahań występują luki systematyczne a drugiego niesystematyczne.

Występowanie systematycznych luk w danych (dziennych o cyklu tygodniowym i miesięcznych o cyklu rocznym), w zależności od ich liczby i układu może pociągać za sobą daleko idące następstwa polegające na tym, że:

- składowe harmoniczne będą ze sobą skorelowane

- niektóre nawet mogą być dokładnie współliniowe (oceny współczynników korelacji będą przyjmować wartości +1 lub -1),
- składowe harmoniczne mogą przyjmować stałe wartości tzn., że ich wariancje będą równe zero,
- niektóre składowe mogą być linio zależne (tworzyć kombinacje liniowe).

Szerszego omówienia wymagają niektóre sytuacje wymienione w powyższych punktach. W przypadku dokładnej współliniowości dwóch lub więcej składowych szacowanych będzie tyle wersji modelu ile składowych jest dokładnie współliniowych. Wersje te będą nierozróżnialne z punktu widzenia ocen parametrów struktury stochastycznej charakteryzujących własności predyktywne (R^2 , SE, V_{SE}) a także innych statystyk (np. DW). Modele te mogą się różnić (i zazwyczaj będą) ocenami błędów prognoz inter- i (lub) ekstrapolacyjnych. Z podobną sytuacją mamy do czynienia wtedy, gdy wystąpi zjawisko liniowej zależności składowych.

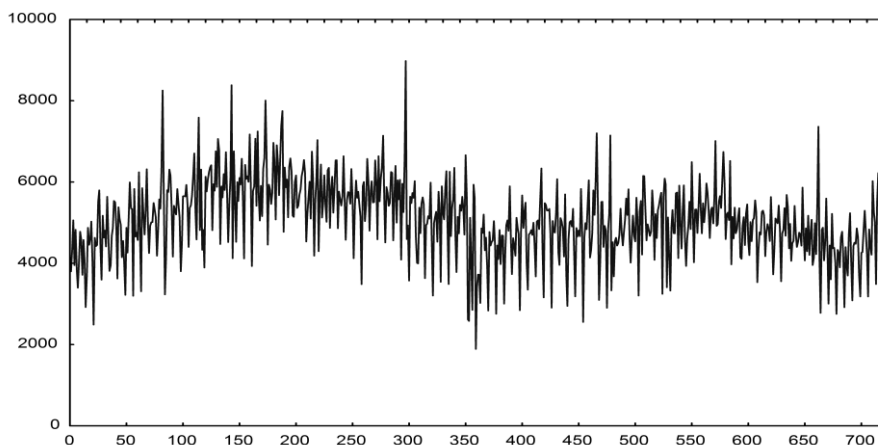
Do stwierdzenia występowania dokładnej współliniowości wystarczy obliczenie współczynników korelacji liniowej. Natomiast wystąpienie zjawiska liniowej zależności składowych harmonicznych nie jest widoczne bezpośrednio w momencie rozpoczęcia modelowania. Uwidacznia się ono w kolejnym etapie zastosowania procedury selekcji *a priori* do wyboru istotnych zmiennych, w tym przypadku składowych harmonicznych w ramach regresji krokowej. Z liniową zależnością będziemy mieć do czynienia wtedy, gdy w kolejnej iteracji otrzymamy co najmniej dwie takie same, co do wartości bezwzględnej, maksymalne oceny współczynników korelacji cząstkowej. Proces estymacji przebiega podobnie jak w przypadku współliniowości tzn. szacuje się tyle wersji modelu ile jest składowych liniowo zależnych. Może się ponadto zdarzyć, że w którejś z następnych iteracji liniową zależność będą tworzyć co najmniej dwie inne składowe. W takim przypadku łączna liczba szacowanych wersji będzie równa iloczynowi liczb składowych liniowo zależnych w kolejnych iteracjach.

W pracy [Szmuksta–Zawadzka, Zawadzki 2013] przedstawiono wykorzystanie modeli szeregu czasowego w prognozowaniu brakujących danych dziennych ze złożoną sezonowością dla luk niesystematycznych. Rozpatrywane były różne warianty występowania tego rodzaju luk równocześnie w określonych dniach i miesiącach. W niniejszej pracy dokonano uogólnienia rozważań na przypadek występowania luk systematycznych.

MODELOWANIE DZIENNEJ SPRZEDAŻY PALIW PŁYNNYCH

Egzemplifikacją przeprowadzonych wyżej rozważań o charakterze teoretycznym będzie przykład empiryczny. Dotyczyć on będzie dziennej sprzedaży paliw płynnych na stacji X w okresie od 7 stycznia pierwszego roku do 31 grudnia roku następnego. Rok trzeci będzie okresem empirycznej weryfikacji prognoz.

Rys. 1. Kształtowanie się dziennej wielkości sprzedaży paliw (w litrach)



Źródło: Szmuksta-Zawadzka i Zawadzki 2011

Rozpatrywany będzie jeden wariant występowania luk systematycznych. Zakładamy, że luki wystąpią w dwóch spośród 12 miesięcy (w sierpniu i październiku) oraz dwóch spośród siedmiu dni tygodnia (środa i piątek). Oznacza to, że luki wystąpią w 294 spośród 724 dni, obejmować więc będą ponad 40% długości szeregu. W szeregu dla pełnych danych składowe harmoniczne dla każdego rodzaju wahań z osobna są ze sobą nieskorelowane. Również nieskorelowane są składowe należące do różnych rodzajów wahań. Natomiast występowanie systematycznych luk w danych sprawiło, że większość bezwzględnych ocen współczynników korelacji składowych harmonicznych o cyklu tygodniowym (13 z 15) przyjęło wartości nie mniejsze od 0,12. Ich oceny mieściły się w przedziale od 0,01 (składowe DC1¹ i DC3) do 0,81 (składowe DS3 i DC1).

Dla składowych o cyklu rocznym w 17 na 54 możliwe kombinacje otrzymano bezwzględne oceny współczynników większe od 0,19. Ocenę najwyższą wynoszącą 0,45 otrzymano dla pary składowych MS1 i MS5. Następnie oceny równe co do wartości bezwzględnej 0,40, ale różniące się znakami, otrzymano dla par składowych: MS2 i MC3 oraz MS4 i MC3.

Występowanie stosunkowo dużej liczby ocen współczynników korelacji dla par składowych różniących się znacznie od zera, może oznaczać, że niektóre z nich mogą być liniowo zależne. W celu sprawdzenia, czy tego rodzaju sytuacja nie ma miejsca, do u wyboru statystycznie istotnych składowych harmonicznych wykorzystana zostanie metoda selekcji *a priori*. W procesie modelowania zostaną także uwzględnione, jako potencjalne zmienne „objaśniające”, zmienne zero-

¹ Symbole DS_j oraz DC_j oznaczają odpowiednio: $D \sin \omega_j, D \cos \omega_j$ a MS_i oraz MC_i: $M \sin \omega_i, M \cos \omega_i$

jedynkowe oznaczające występowanie świąt oraz dni około świątecznych (tzn. dnia poprzedzającego święto oraz dnia po nim następującego).

W Tabeli 1 zestawione zostały oceny współczynników korelacji cząstkowej otrzymane po 15 iteracji dla składowych o maksymalnych, co do wartości bezwzględnej, ich ocenach.

Tabela 1. Zmienne poza równaniem po 15 iteracji o maksymalnych ocenach współczynników korelacji cząstkowej

		Cząstk.	Semicz.		Minimum		
Zmienne	Beta dla	Kor.	Kor.	Teleran.	Toleran.	t(413)	poziom p
DS3	-0,3887	-0,1053	-0,0750	0,0372	0,0372	-2,1524	0,0319
DC2	0,1228	0,1053	0,0750	0,3724	0,3724	2,1524	0,0319
DC3	0,0811	0,1053	0,0750	0,8546	0,7412	2,1524	0,0319

Źródło: obliczenia własne

Oceny te wynoszące 0,1053, otrzymano dla składowych harmonicznych: DS3, DC2 i DC3. Dla pierwszej składowej otrzymano ocenę ujemną a dwóch pozostałych dodatnie.

Otrzymanie identycznych bezwzględnych ocen współczynników korelacji cząstkowej wskazuje, że składowe DS3, DC2 i DC3 są liniowo zależne. Oznacza to konieczność szacowania trzech wersji modelu, z których każda zawierać będzie tylko jedną z tych składowych.

Z kolei w Tabeli 2 zestawione zostały zmienne (składowe) charakteryzujące się maksymalnymi ocenami współczynników korelacji cząstkowej po 21 iteracji dla zmiennych, które dotąd nie weszły do modelu.

Tabela 2. Zmienne poza równaniem po 21 iteracji o maksymalnych ocenach współczynników korelacji cząstkowej

		Cząstk.	Semicz.		Minimum		
	Beta dla	Kor.	Kor.	Teleran.	Toleran.	t(407)	poziom p
MS2	-0,0411	-0,0459	-0,0321	0,6092	0,0370	-0,9272	0,3544
MS4	-0,0411	-0,0459	-0,0321	0,6090	0,0370	-0,9272	0,3544
DC2	-	-	-	0,0000	-	-	-
DC3	-	-	-	0,0000	-	-	-

Źródło: obliczenia własne

Z tabeli tej wynika, że maksymalne, co do wartości bezwzględnej, oceny współczynników korelacji cząstkowej otrzymano dla składowych: MS2 i MS4. Oznacza to, podobnie jak poprzednio, konieczność szacowania równań zawierających jedną z nich. Brak charakterystyk statystycznych dla zmiennych (składowych) DC2 i DC3 świadczy o tym, że tworzą one kombinację liniową i są

liniowo zależne. Z uwagi na to, że po 15-tej iteracji otrzymano identyczne oceny współczynników korelacji cząstkowej dla trzech składowych a po 21-szej dla dwóch, trzeba będzie szacować sześć równań zawierających po jednej składowej z każdej grupy. Równania te zawierać będą oczywiście te same pozostałe zmienne wybrane w wcześniejszych iteracjach.

W Tabeli 3 zestawione zostały oszacowania wybranych parametrów struktury stochastycznej, będące syntetycznymi miernikami ich własności predykcyjnych wyrażonych za pomocą współczynników determinacji (R^2) oraz odchyłeń standardowych składników losowych (SE) i współczynników zmienności losowej (V_{SE}) dla sześciu wersji modeli.

Tabela 3. Wyniki oszacowań modeli oszczędnych dla systematycznych luk danych

	DS3_MS2	DS3_MS4	DC2_MS2	DC2_MS4	DC3_MS2	DC3_MS4
R^2	0,4869	0,4869	0,4869	0,4869	0,4869	0,4869
SE	690,01	690,01	690,01	690,01	690,01	690,01
V_{SE}	0,1399	0,1399	0,1399	0,1399	0,1399	0,1399

Źródło: obliczenia własne

Równania te oznaczono podwójnymi symbolami oddzielonymi podkreślnikami, przy czym pierwszy symbol odnosi się do jednej z trzech składowych liniowo zależnych o cyklu tygodniowym (DS3, DC2 lub DC3) a drugi dla składowych zależnych liniowo o cyklu rocznym (MS2 lub MS4).

Z uwagi na to, że oceny syntetycznych mierników własności predykcyjnych są takie same, równania te są nierozróżnialne z tego punktu widzenia. W kolejnych dwóch wierszach zamieszczone zostały oceny błędów ogółem prognoz interpolacyjnych i ekstrapolacyjnych. Zgodnie z tym co powiedziano w poprzednim podrozdziale, mogą się one i zazwyczaj będą różnić ocenami błędów prognoz inter- i (lub) ekstrapolacyjnych.

WYNIKI PROGNOZOWANIA INTER- I EKSTRAPOLACYJNEGO

Na podstawie oszacowanych równań dla dni i miesięcy, w których wystąpiły luki zostały wyznaczone prognozy interpolacyjne. Z uwagi na to, że luki otrzymano przez „wymazanie” części danych z pełnego szeregu, możliwe było obliczenie średnich względnych błędów tych prognoz. W wierszu drugim Tabeli 4 zamieszczone zostały oceny błędów prognoz interpolacyjnych ogółem obliczone dla 294 dni. Pozostałe wiersze zawierają ich dekompozycję dla tych dni tygodnia i miesięcy, w których wystąpiły luki. W celach porównawczych podane zostały uśrednione oceny względnych błędów prognoz dla pozostałych dni i miesięcy nie zawierających luk.

Tabela 4. Oceny błędów prognoz interpolacyjnych

	DS3_MS2	DS3_MS4	DC2_MS2	DC2_MS4	DC3_MS2	DC3_MS4
Ogółem	0,2130	0,2186	0,1133	0,1189	0,1031	0,1087
Środa	>0,3000	>0,3000	0,1339	0,1346	0,1127	0,1146
Piątek	0,2114	0,2127	0,1172	0,1194	0,1093	0,1104
Poz. dni	0,0836	0,0989	0,0836	0,0989	0,0836	0,0989
Sierpień	0,1118	0,1188	0,0959	0,1129	0,0904	0,1080
Październik	0,1483	0,1679	0,1011	0,1107	0,0954	0,1045
Poz. m-ce	0,2736	0,2736	0,1244	0,1244	0,1100	0,1100

Źródło: obliczenia własne

Spośród sześciu oszacowanych modeli zdecydowanie niższymi ocenami błędów prognoz interpolacyjnych ogółem charakteryzują się modele z dodatnimi ocenami współczynników korelacji cząstkowej. Oceny błędów zawarte są w przedziale od 10,31% (DC3_MS24) do 11,89% (DC2_MS4). Natomiast znacznie wyższe oceny błędów prognoz, przynajmniej o 9 punktów procentowych, otrzymano na podstawie modeli o ujemnej ocenie współczynnika korelacji cząstkowej (DS3_MS2, DS3_MS4).

Obecnie analizie poddane zostaną oceny błędów dla dni i miesięcy z lukami. Minimalną ocenę błędu dla środy, wynoszącą 11,27%, otrzymano dla predyktora DC3_MS2 i tylko nieznacznie wyższą dla predyktora C3_MS3. Oceny błędów prognoz z pierwszą składową DC2 były wyższe o ok. 2 punkty procentowe. Zdecydowanie najwyższe oceny błędów, przekraczające 30 procent, otrzymano dla predyktorów o ujemnej ocenie współczynnika korelacji cząstkowej. W przypadku piątku najlepszym okazał się także predyktor DC3_MS2. Oceny te dla pozostałych trzech predyktorów o dodatnich ocenach współczynników korelacji cząstkowej były wyższe nie więcej niż o 1 punkt procentowy. Natomiast najwyższe oceny błędów, podobnie jak dla środy, otrzymano dla predyktorów o ujemnej ocenie tego współczynnika. Uśrednione oceny błędów dla pozostałych dni przyjęły wartość 8,36% dla predyktorów z drugą składową MS2 oraz 9,89% ze składową MS3.

Ocenę minimalną dla sierpnia, wynoszącą 9,04%, otrzymano także dla predyktora DC3_MS2. Oceny błędów prognoz otrzymanych na podstawie pozostałych modeli były nie więcej niż o 2,89 punktu wyższe, przy czym dotyczy to także modeli DS3_MS2 oraz DS3_MS4. Oceny błędów otrzymanych na podstawie predyktorów o dodatnich ocenach współczynników korelacji cząstkowej zawarte były w przedziale od 9,54% (DC3_MS2) do 11,17% (DC2_MWS4). Natomiast dla predyktorów o ujemnej ocenie tego współczynnika wynosiły one odpowiednio: 14,83% i 16,79%. Uśrednione oceny otrzymane dla tych

predyktorów wynosiły 27,36%. Dla predyktorów z pierwszą składową DC3 kształtowały się one na poziomie 11,00%, a ze składową DC2 12,44%.

Na podstawie oszacowanych równań zostały wyznaczone także prognozy ekstrapolacyjne a następnie przeprowadzona ich empiryczna weryfikacja tzn. obliczone zostały przeciętne względne błędy prognoz. Było to możliwe, ponieważ dla okresu empirycznej weryfikacji prognoz znane były realizacje zmiennej prognozowanej.

W Tabeli 5 zestawione zostały oceny błędów prognoz ekstrapolacyjnych ogółem oraz błędy prognoz zdezagregowanych na dni tygodnia i miesiące w których wystąpiły luki.

Z informacji zawartych w drugim wierszu wynika, że oceny błędów prognoz ekstrapolacyjnych ogółem są o ok. 5-6 punktów procentowych wyższe od ocen błędów prognoz interpolacyjnych, otrzymanych na podstawie predyktorów z dodatnimi ocenami współczynników korelacji cząstkowej. Natomiast są one o ok. 3 punkty niższe od ocen błędów prognoz interpolacyjnych otrzymanych na podstawie predyktorów z ujemną oceną tego współczynnika. Najniższą ocenę wynoszącą 15,94% charakteryzuje się predyktor DC3_MS2. Dla pozostałych trzech predyktorów z dodatnimi ocenami współczynników korelacji cząstkowej są one wyższe nie więcej niż o 0,57 punktu procentowego.

Tabela 5. Oceny błędów prognoz ekstrapolacyjnych

	DS3_MS2	DS3_MS4	DC2_MS2	DC2_MS4	DC3_MS2	DC3_MS4
Ogółem	0,1848	0,1877	0,1626	0,1651	0,1595	0,1618
Środa	0,2975	0,2978	0,1925	0,1922	0,1319	0,1353
Piątek	0,1717	0,1777	0,1202	0,1235	0,1583	0,1565
Poz. dni	0,1651	0,1680	0,1651	0,1680	0,1651	0,1680
Sierpień	0,1624	0,1897	0,1738	0,2152	0,1716	0,2130
Październik	0,2065	0,2123	0,1741	0,1607	0,1749	0,1590
Poz. m-ce	0,1874	0,1874	0,1612	0,1612	0,1528	0,1528

Źródło: obliczenia własne

Oceny wyższe o ok. 2-3 punkty procentowe otrzymano dla predyktorów z ujemnymi ocenami współczynników korelacji cząstkowej (DS3_MS2 i DS3_MS4). W wierszach od 3 do 5 zestawiono oceny błędów prognoz dla dni tygodnia z lukami i bez luk. W przypadku środy widoczne jest bardzo duże ich zróżnicowanie dla par predyktorów, różniących się pierwszą składową. Zdecydowanie najniższe oceny, kształtujące się w granicach 13 procent otrzymano dla predyktorów: DC3_MS2 oraz DC3_MS4. W przypadku drugiej pary z dodatnią oceną współczynnika korelacji cząstkowej (MC2_MS2 oraz MC2_MS4) oceny te są o około 6 punktów procentowych wyższe. Najwyższe oceny błędów, wynoszące ok. 30 procent otrzymano dla pary predyktorów

z ujemną oceną współczynnika korelacji cząstkowej (DS3_MS2 oraz DS3_MS4). Najniższe oceny dla piątku, wynoszące odpowiednio 12,01% oraz 12,34% otrzymano dla pary predyktorów z pierwszą składową DC2. W przypadku pary z pierwszą składową DC3 oceny te są o około 3 punkty procentowe wyższe. Oceny błędów otrzymane dla predyktorów o ujemnej ocenie współczynnika korelacji cząstkowej (z pierwszą składową DS3) są o dalsze 1-2 punkty procentowe wyższe. Uśrednione oceny błędów dla pozostałych dni tygodnia przyjęły wartość 16,51% dla predyktorów z drugą składową MS2 i 16,80% ze składową MS4.

Kolejne trzy wiersze zawierają oceny błędów prognoz dla miesięcy z lukami i uśrednione dla miesięcy bez luk. Dla sierpnia niższe oceny błędów dla par predyktorów różniących się pierwszą składową otrzymano dla tych, w których występowała druga składowa MS2. Ich oceny zawarte są w przedziale od 16,15% (dla predyktora DS3_MS2) do 17,40% dla predyktora DC2_DS2). Oceny błędów otrzymane dla predyktorów z drugą składową są w ramach poszczególnych par są o ok. 2,5 – 4 procent wyższe. Dla października widoczny był wpływ znaku oceny współczynników korelacji cząstkowej. Dla predyktorów z dodatnią jego oceną, błędy kształtowały się na poziomie od 15,90 % (DC3_MS4) do 17,49% (DC3_MS2). Dla predyktorów z oceną ujemną, przekraczają one 20 procent. Dla pozostałych miesięcy oceny błędów otrzymanych dla par predyktorów są takie same. Kształtują się one w przedziale od 15,28% dla predyktorów z pierwszą składową DC3 do 18,74% ze składową DS3.

WNIOSKI

Z przeprowadzonych w pracy rozważań można wyprowadzić następujące wnioski:

1. W procesie wyboru statystycznie istotnych składowych harmonicznych potwierdzona została zasadność stosowania procedury *selekcji a priori* w modelowaniu w warunkach występowania systematycznych luk w danych.
2. Jej zastosowanie pozwala zidentyfikować liniowo zależne składowe harmoniczne. Tworzą ją, dla przynajmniej jednej z kolejnych iteracji, składowe o maksymalnych bezwzględnych ocenach współczynników korelacji cząstkowej. Oznacza to konieczność szacowania liczby wersji modelu będącej iloczynem liczb: zmiennych liniowo zależnych i iteracji.
3. Oszacowane wersje charakteryzują się identycznymi ocenami syntetycznych ocen własności predykcyjnych (R^2 , Se , V_{Se}). Natomiast różnią się ocenami błędów prognoz inter- i ekstrapolacyjnych.
4. Z przykładu empirycznego, obejmującego występowanie luk systematycznych w dwóch wybranych dniach tygodnia i dwóch miesiącach, wynika, że minimalne oceny błędów ogółem prognoz interpolacyjnych są niższe niż prognoz ekstrapolacyjnych.
5. Istotnym czynnikiem determinującym dokładność prognoz okazał ujemny znak współczynnika korelacji cząstkowej dla jednej z liniowo zależnych składowych

o cyklu tygodniowym (DS3). Oceny błędów ogółem prognoz interpolacyjnych były blisko dwukrotnie wyższe od błędów otrzymanych dla predyktorów z dodatnimi jego wartościami. Dla prognoz ekstrapolacyjnych różnice w dokładności wynosiły 2-3 punkty procentowe.

Z przedstawionych wyżej wniosków wynika, że oszczędne modele harmoniczne szeregu czasowego ze złożoną sezonowością dla danych dziennych, w których wahania o cyklu rocznym i tygodniowym opisywanych za pomocą wielomianów trygonometrycznych, mogą być z powodzeniem wykorzystywane w prognozowaniu w warunkach występowania luk systematycznych obejmujących nawet ok. 40 % długości szeregu.

BIBLIOGRAFIA

- Kufel T. (2010) Ekonometryczna analiza cykliczności procesów gospodarczych o wysokiej częstotliwości obserwowania, Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, Toruń.
- Szmuksta-Zawadzka M., Zawadzki J. (2011) The ex post analysis of the forecast accuracy of liquid fuel sales for daily data, „Econometrics”, Research Papers of Wrocław University of Economics Nr 32, s.194-204, Wrocław.
- Szmuksta-Zawadzka M., Zawadzki J. (2013) Wykorzystanie oszczędnych modeli harmonicznych z cyklicznością w prognozowaniu brakujących danych(w druku)

HARMONICAL MODELS WITH COMPLEX SEASONALITY IN FORECASTING TIME SERIES WITH SYSTEMATIC GAPS

Abstract: In the modeling of the variables with complex seasonality for complete time series and with unsystematic data gaps can be used both types of models: with dummy variables and harmonic models. However, in modeling variable with systematic gaps can be used only harmonic models. In these models, each type of fluctuation is described by separate sets of sine- and cosine component. Theoretical considerations are illustrated by an empirical example.

Keywords: harmonic models, complex seasonality, missing data

WIELOKRYTERIALNY MODEL OPTYMALIZACYJNY PRZECIĘTNEGO GOSPODARSTWA ROLNEGO W WOJEWÓDZTWIE ZACHODNIOPOMORSKIM

Jadwiga Zaród

Katedra Zastosowań Matematyki w Ekonomii
Zachodniopomorski Uniwersytet Technologiczny w Szczecinie
e-mail: jzarod@zut.edu.pl

Streszczenie: Na podstawie danych statystycznych zbudowano wielokryterialny model optymalizacyjny. Jego funkcje celu maksymalizują dochód rolniczy i produkcję rolniczą oraz minimalizują straty substancji organicznej w glebie. Do rozwiązania modelu wykorzystano optymalizację celową. Wyniki rozwiązania wskazały taką strukturę produkcji, która: dała najwyższy w danych warunkach dochód rolniczy, dostarczyła produkcji, o wysokim standardzie na potrzeby własne i na sprzedaż oraz zapobiegła degradacji środowiska naturalnego w dłuższym okresie czasu.

Celem tego artykułu jest badanie zrównoważonego rozwoju produkcji rolnej na przykładzie przeciętnego gospodarstwa rolnego w województwie zachodniopomorskim. Istotą zrównoważonego rozwoju jest zachowanie równowagi pomiędzy zadaniami ekonomicznymi, produkcyjnymi i ekologicznymi.

Słowa kluczowe: wielokryterialny model optymalizacyjny, dochód rolniczy, produkcja rolnicza

WSTĘP

Rolnictwo to dział gospodarki narodowej ściśle związany z środowiskiem naturalnym. Jego zrównoważony rozwój polega na racjonalnym wykorzystaniu zasobów przyrody i ograniczaniu negatywnych wpływów na środowisko. Wiąże się to ze stosowaniem odpowiedniego następstwa i doboru roślin, prawidłowego wykonywania zabiegów agrotechnicznych oraz umiejętnego nawożenia. Działalność rolnicza ukierunkowana jest na wytwarzanie produkcji roślinnej i zwierzęcej w zamian za uzyskany dochód. Natomiast istotą rolnictwa

zrównoważonego jest zachowanie równowagi pomiędzy celami produkcyjnymi, ekonomicznymi i ekologicznymi [Staniak 2009]. Cel produkcyjny jest realizowany poprzez wytwarzanie produkcji o właściwościach wymaganych przez konsumentów lub przemysł przetwórczy. Działania ekonomiczne skupiają się na wypracowaniu dochodu zapewniającego rolnikom godziwy poziom życia i umożliwiającego rozwój gospodarstw rolnych. Zamierzenia ekologiczne dążą do zagwarantowania stabilności agrosystemu i zapobiegania degradacji środowiska przyrodniczego.

Celem tego artykułu jest badanie zrównoważonego rozwoju produkcji w przeciętnym gospodarstwie rolnym województwa zachodniopomorskiego w latach 2008 – 2011. Do analizy tego zagadnienia wykorzystano wielokryterialny model optymalizacyjny. Jego funkcje celu maksymalizują dochód rolniczy i wielkość produkcji oraz minimalizują straty substancji organicznych w glebie. W wyniku rozwiązania optymalnego zostanie ustalona taka struktura zasiewów, która zapewni produkcję towarową i na użytek własny o wysokim standardzie, da najwyższy, możliwy w danych warunkach dochód rolniczy oraz nie zdegraduje środowiska naturalnego.

METODA BADAWCZA

Podstawową metodą badawczą pracy jest liniowo – dynamiczny model optymalizacyjny z trzema kryteriami celu. Matematyczny zapis takiego modelu [Krawiec 1991] ma postać:

$$ax(t) \leq b(t) \quad \text{warunki ograniczające} \quad (1)$$

$$x(t + 1) \leq x(t) + f_t[x(t), u(t)] \quad \text{równania dynamiki} \quad (2)$$

$$F = \max \{F_1, F_2, F_3\} \quad \text{kryteria sterowania} \quad (3)$$

$$x(t) \geq 0, \quad u(t) \geq 0 \quad \text{warunki brzegowe} \quad (4)$$

gdzie : t - stany (kolejne lata gospodarowania), $t = 0, 1, 2, \dots, k$.

a – parametry techniczno-ekonomiczne,

$b(t)$ – wektor ograniczeń w kolejnych stanach

$x(t)$ – wektor stanu,

$u(t)$ – wektor sterowań,

Należy przyjąć, że początkowy stan układu w momencie $t=0$ jest znany i opisuje areal roślin i stan zwierząt w momencie poprzedzającym pierwszy rok badań.

Wektor stanu $x(t)$ przyjmuje postać:

$$x(t) = [x_t(t), x_p(t), x_r(t)] = [x_1(t), \dots, x_n(t)] \quad (5)$$

gdzie : $x_t(t)$ – wektor stanu działalności towarowych (opisuje areal roślin towarowych uprawianych w roku t , liczbę zwierząt odpowiednich klas i gatunków, które w roku t dają produkcję towarową, np. mleko, mięso);

$x_p(t)$ – wektor stanu działalności nietowarowych (opisuje areał roślin pastewnych uprawianych w roku t oraz klasy i gatunki zwierząt nie dających produkcji towarowej);

$x_r(t)$ – wektor stanu pozostałych działalności, np. zakup środków produkcji, pasz.

Wektor sterowań $u(t) = u_{ij}(t)$ przedstawia przepływy wewnątrz gospodarstwa lub między gospodarstwem a otoczeniem. Składowe tego wektora opisują areały następujących po sobie roślin, zmianę klas zwierząt, sprzedaż lub zakup zwierząt, zakup lub sprzedaż gruntów ornych, przy przejściu gospodarstwa ze stanu t do $t+1$. Wskaźniki i, j określają porządek następowania po sobie np. po roślinie i będzie uprawiana roślina j lub zwierzę z klasy i przejdzie do klasy j .

Równania dynamiki dla produkcji roślinnej przyjmują postać:

$$x_i(t+1) = \sum_p u_{pi}(t) \quad (6)$$

gdzie: $x_i(t+1)$ – areał i -tej rośliny uprawianej w roku $t+1$,

$u_{pi}(t)$ – areał różnych przedplonów p , po których jest uprawiana i -ta roślina w roku $t+1$.

Postać równań dynamiki dla produkcji zwierzęcej jest następująca:

$$x_i(t+1) = x_i(t) - u_{is}(t) + u_{iz}(t) + x_{ji}(t) \quad (7)$$

gdzie: $x_i(t+1)$ – stan zwierząt i -tego gatunku w roku $t+1$,

$x_i(t)$ – stan zwierząt i -tego gatunku w roku poprzednim,

$u_{is}(t)$ – sprzedaż zwierząt i -tego gatunku w roku t ,

$u_{iz}(t)$ – zakup zwierząt i -tego gatunku w roku t ,

$u_{ji}(t)$ – liczba zwierząt i -tego gatunku z własnej hodowli, z przeklasowania.

Oprócz równań dynamiki we wzorze (2) występują ograniczenia stanu, które dotyczą powierzchni gruntów ornych i użytków zielonych oraz stanowisk dla inwentarza. Podsumowując, należy powiedzieć, że liniowe równania dynamiki i ograniczenia przeprowadzają gospodarstwo rolne ze stanu t do $t+1$, czyli z poprzedniego do następnego roku badań. Uwzględniają: stan w jakim znajduje się gospodarstwo w roku t , sterowania, jakie można stosować przy przejściu ze stanu t do $t+1$ oraz ograniczenia jakim podlega produkcja rolnicza.

Kryterium celu F_1 dotyczy dochodu rolniczego brutto i wyraża się wzorem:

$$F_1 = \sum_t [m(t)^T u(t) + w(t+1)^T x(t+1)] \rightarrow \max \quad (8)$$

Wektor $m(t) = [m_{jp}(t)]$ określa jednostkowy dochód rolniczy dla sterowań $u_{pi}(t)$ prowadzących do działalności towarowych lub wielkość nakładów jednostkowych ponoszonych przy stosowaniu sterowań $u_{pi}(t)$ w przypadku działalności nietowarowych. Natomiast składowe wektora $w(t+1) = [w_1(t+1), \dots, w_n(t+1)]$ opisują ceny produktów towarowych lub środków produkcji (pasz, nawozów) zmiennych stanu $w_i(t+1)$ oznaczających wielkość sprzedaży lub zakupu.

F_2 to kryterium sterowania maksymalizujące wielkość produkcji o postaci:

$$F_2 = \sum_t [g(t)^T u(t) + k(t+1)^T x(t+1)] \rightarrow \max \quad (9)$$

gdzie: $g(t)$, $k(t+1)$ – wektor jednostkowych wydajności zmiennych sterowań i stanu w kolejnych latach;

Funkcja F_3 maksymalizuje ilość substancji organicznej w glebie:

$$F_3 = \sum_t [o(t)^T u(t) + p(t+1)^T x(t+1)] \rightarrow \max \quad (10)$$

gdzie: $o(t)$, $p(t+1)$ – wektor jednostkowych współczynników reprodukcji lub degradacji gleby dla zmiennych stanu i sterowań.

Istnieje kilka metod rozwiązywania wielokryterialnych modeli optymalizacyjnych. Najbardziej popularnym sposobem jest programowanie celowe. Jego twórcami są Charnes i Cooper [Charnes, Cooper 1961]. W podejściu tym należy rozwiązać zbudowany model oddzielnie ze względu na każde kryterium. Po uzyskaniu optymalnych wyników ze względu na kryteria sterowania, każdą funkcję celu traktuje się jako kolejny warunek ograniczający modelu o postaci:

$$m(t)^T u(t) + w(t+1)^T x(t+1) = dr \quad (11)$$

$$g(t)^T u(t) + k(t+1)^T x(t+1) = pr \quad (12)$$

$$o(t)^T u(t) + p(t+1)^T x(t+1) = so \quad (13)$$

gdzie: dr – najwyższa wartość dochodu rolniczego uzyskana w rozwiązaniu modelu jednokryterialnego,

pr – optymalna wielkość produkcji rolniczej uzyskanej w rozwiązaniu modelu jednokryterialnego,

so – ilość substancji organicznej zachowanej w glebie w wyniku optymalnego rozwiązania modelu jednokryterialnego.

We wszystkich tych warunkach występuje restrykcyjne ograniczenie typu równości, które należy osłabić. Pełnym osłabieniem równości nazywamy przekształcenie, w którym wystąpią zmienne niedoboru (u^-) lub nadmiaru (u^+) wyrażające wielkości niespełnienia osiągniętych wartości w modelach jednokryterialnych [Szapiro 2001]. Po przekształceniu dodane warunki ograniczające przyjmą postać:

$$m(t)^T u(t) + w(t+1)^T x(t+1) - u_1^+ + u_2^- = dr \quad (14)$$

$$g(t)^T u(t) + k(t+1)^T x(t+1) - u_3^+ + u_4^- = pr \quad (15)$$

$$o(t)^T u(t) + p(t+1)^T x(t+1) - u_5^+ + u_6^- = so \quad (16)$$

Następnie zastępuje się wiele kryteriów jedną funkcją dystansową, opisującą koszty (kary) odchylenia od wartości docelowych. W funkcji tej wystąpią obie zmienne, dotyczące nadmiaru lub niedoboru dochodu rolniczego i produkcji rolniczej, ponieważ nie zakłada się szczególnych zaleceń co do sposobu ich osiągnięcia. Natomiast niedobór substancji organicznej w glebie należy

minimalizować, aby nie zdegradować środowiska naturalnego. Funkcja dystansowa będzie miała postać:

$$F = u_1^+ + u_2^- + u_3^+ + u_4^- + u_6^- \rightarrow \min \quad (17)$$

MATERIAŁ BADAWCZY

Do badania wykorzystano dane statystyczne dotyczące gospodarstw rolnych w latach 2008-2011. Niektóre z tych danych dla przeciętnego gospodarstwa rolnego zawiera tabela 1.

Tabela 1. Podstawowe charakterystyki gospodarstw rolnych

Wyszczególnienie	Statystyki w latach			
	2008	2009	2010	2011
Powierzchnia gospodarstwa (ha)	17,71	18,62	23,23	20,50
Powierzchnia UR (ha)	16,19	16,87	19,88	18,66
Powierzchnia zasiewów (ha)	12,25	12,75	14,10	13,56
Powierzchnia TUZ (ha)	2,36	2,55	3,15	3,03
Struktura zasiewów (%)	100	100	100	100
w tym: zboża	72,1	73	64,3	73,92
rzepak	14,4	15,1	17,7	14,84
rośliny okopowe	4,8	5,0	4,4	4,2
pastewne	6,0	4,6	6,5	2,7
inne uprawy	2,7	2,3	7,1	4,34
Plony (dt·ha ⁻¹):				
zboż	30,8	41,2	44,9	35,3
rzepaku	27,3	33,0	25,7	17,2
ziemniaków	215	277	232	254
buraków	400	460	459	559
łąk	30,7	36,5	36,4	38,9
pastwisk	121	215	149	177
Ceny (zł·dt ⁻¹):				
zboż	55,31	41,84	45,78	70,49
rzepaku	128,60	112,54	123,94	174,46
ziemniaków	23,94	23,92	24,69	28,13
buraków	10,49	12,56	11,60	14,41
cukrowych				
Nawożenie NPK (kg)	130,1	121,4	122,9	124,9
Liczba krów w gospodarstwie	3	3	3	3
Liczba loch w gospodarstwie	2	2	2	2
Dotacje bezpośrednie (zł·ha ⁻¹)	608,63	863,45	889,34	984,79

Źródło: opracowanie własne na podstawie danych GUS

Informacje dotyczące przeciętnego gospodarstwa rolnego w analizowanych latach posłużyły do ustalenia parametrów techniczno-ekonomicznych, wyrazów wolnych oraz współczynników funkcji celów wielokryterialnego modelu optymalizacyjnego. Zbudowany model składał się ze 100 zmiennych stanu

i sterowań oraz 122 warunków bilansowych. Zapewniał dobre pokrycie gleb roślinnością, terminowe wykonanie zabiegów agrotechnicznych, niskie dawki nawozów mineralnych oraz tylko niezbędne chemiczne zabiegi ochrony roślin. Równania dynamiki uwzględniały następstwo roślin na poszczególnych polach w badanym okresie (Tabela 2) oraz zamknięty obrót stada zwierząt. Powierzchnię gruntów ornych przyjęto na poziomie wartości średniej z czterech lat (2008-2011). Liczbę krów i loch ustalono tylko na podstawie gospodarstw zajmujących się produkcją zwierzęcą.

Tabela 2. Plodozmiian i zmianowanie roślin uwzględnione w modelu

Pola Lata	Pole I	Pole II	Pole III	Pole IV
2008	buraki ziemniaki owies	pszenica jęczmień	pszenżyto rzepak	żyto pozostałe uprawy
2009	pszenica jęczmień	pszenżyto rzepak	żyto pozostałe uprawy	buraki ziemniaki owies
2010	pszenżyto rzepak	żyto pozostałe uprawy	buraki ziemniaki owies	pszenica jęczmień
2011	żyto pozostałe uprawy	buraki ziemniaki owies	pszenica jęczmień	pszenżyto rzepak

Źródło: opracowanie własne

Jednostkowe dochody rolnicze (parametry pierwszej funkcji celu) stanowiły różnicę pomiędzy wartością produkcji (np. cena x plon) a jej kosztami [Augustyńska–Grzymek, 2012]. Kalkulacje kosztów poszczególnych upraw i gatunków zwierząt oszacowano na podstawie opracowań ZODR [Kalkulacje rolnicze, 2008, 2009, 2010, 2011]. Uzyskane dochody powiększono o dotacje bezpośrednie (Jednolite Płatności Obszarowe i dopłaty uzupełniające) a w przypadku buraków cukrowych jeszcze o płatności cukrowe. Zasiwy przeznaczone na paszę oraz zwierzęta do odchowu w funkcji celu obarczono kosztami pomniejszonymi o dopłaty a nawozy mineralne i mieszanki paszowe jednostkową ceną zakupu.

Plony oraz produkty pochodzenia zwierzęcego to współczynniki drugiej funkcji celu. W celach porównawczych w modelu zostały one uwzględnione wartościowo (w zł), z pominięciem dotacji, gdyż w innych jednostkach wyraża się produkcję roślinną (w dt) a w innych produkcję zwierzęcą (w kg).

Do wyznaczania parametrów trzeciej funkcji celu wykorzystano współczynniki reprodukcji i degradacji glebowej substancji organicznej według Eicha i Kindlera (Tabela 3).

Tabela 3. Współczynniki reprodukcji i degradacji materii organicznej gleby

Rośliny lub nawóz organiczny	Współczynniki reprodukcji (+) i degradacji (-) dla gleb w tonach materii organicznej na ha			
	gleby			
	lekkie	średnie	ciężkie	czarne ziemie
Okopowe	- 1,26	-1,40	-1,54	-1,02
Kukurydza	-1,12	-1,15	-1,22	-0,91
Zboża i oleiste	-0,49	-0,53	-0,56	-0,38
Poplony ozime i zboża na zielonkę	-0,32	-0,35	-0,38	-0,25
Strączkowe	+0,32	+0,35	+0,38	+0,38
Wsiewki motylkowe i trawy	+0,63	+0,70	+0,77	+0,77
Trawy	+0,95	+1,05	+1,16	+1,16
Motylkowe	+0,89	+1,96	+2,10	+2,10
Obornik*	+0,35		(ok.4-5t świeżej masy)	
Gnojowica*	+0,28		(ok.10-16t świeżej masy)	
Słoma*	+0,21		(ok.1,1t świeżej masy)	

Źródło: opracowanie na podstawie [Fotyła, Mercik 1992]

Objaśnienia: *na 1tonę suchej masy nawozu

Wartości współczynników degradacji i reprodukcji określają stopień zubożenia lub wzbogacenia gleby w substancję organiczną w t·ha⁻¹ przy uprawie danego gatunku roślin lub zastosowaniu określonej dawki nawozów organicznych.

WYNIKI BADAŃ

Rozwiązanie liniowo-dynamicznego modelu optymalizacji wielokryterialnej przebiegało dwuetapowo. W pierwszym etapie rozwiązano model optymalizujący każde kryterium oddzielnie. Z rozwiązań modeli jednokryterialnych wynika, że w przeciętnym gospodarstwie rolnym województwa zachodniopomorskiego istniała możliwość uzyskania w latach 2008 - 2011 dochodu rolniczego w wysokości 168479,19zł, wartości produkcji równej 241979,60zł oraz 2,30 t substancji organicznych w glebie. W drugim etapie dopisano do modelu 6 zmiennych osłabiających równości oraz 3 warunki elastyczne i rozwiązano model z funkcją dystansową. Rozwiązaniem tak postawionego problemu jest zestaw zmiennych decyzyjnych tworzących strukturę produkcji (tabela 4), której realizacja zapewni najwyższy, w danych warunkach, dochód rolniczy i produkcję rolniczą oraz dodatni bilans materii organicznej w glebie.

Tabela 4. Wyniki rozwiązania modelu wielokryterialnego

Wyszczególnienie	2008	2009	2010	2011
Grunty orne (ha)	13,17	13,17	13,17	13,17
Pszenica (ha)	3,42	2,37	3,56	3,82
Jęczmień (ha)	-	-	-	-
Żyto (ha)	3,03	3,42	2,50	1,71
Owies (ha)	1,05	2,37	2,37	2,23
Pszennyto (ha)	1,98	1,44	-	1,98
Rzepak (ha)	1,84	1,98	2,37	1,58
Ziemniaki	0,40	0,27	0,40	0,05
Buraki cukrowe (ha)	0,70	0,73	0,84	0,59
Buraki pastewne (ha)	0,22	0,19	0,21	0,20
Pozostałe uprawy (ha)	0,53	0,4	0,92	0,66
Łąki (ha)	2,13	1,4	1,65	1,53
Pastwiska (ha)	1,34	0,79	0,80	0,63
Krowy (szt.)	3	3	3	3
Cielęta (szt.)	2,94	2,94	2,94	2,94
Młode bydło opasowe (szt.)	2	2,35	2,35	2,35
Jałówki remontowe (szt.)	0,38	0,38	0,38	0,38
Krowy wybrakowane (szt.)	0,38	0,38	0,38	0,38
Lochy (szt.)	2	2	2	2
Prosięta (szt.)	32	32	32	32
Tuczniaki (szt.)	32	31,04	31,04	31,04
Loszki remontowe (szt.)	-	0,4	0,4	0,4
Lochy wybrakowane (szt.)	0,4	0,4	0,4	0,4
Dochód rolniczy (zł)	161650,02			
Produkcja rolnicza	232075,31			
Substancja organiczna w glebie (t)	0,57			

Źródło: obliczenia własne

Uzyskane rozwiązanie optymalne jest zgodne z zasadami zrównoważonego rolnictwa na poziomie gospodarstwa rolnego [Kuś, Krasowicz 2001]. Model nie opisuje walorów jakościowych produktów rolnych, ale stosowane zabiegi agrotechniczne wskazują na ich wysoki standard. Z oceny sanitarnej wynika, że w województwie zachodniopomorskim w 2010 r. zdyskwalifikowano tylko 2% ogółu zbadanych próbek warzyw, 1,1% próbek mięsa i przetworów mięsnych oraz 2,2% próbek ziarna i przetworów zbożowych. Mniej korzystnie wypadła ocena mleka i jego przetworów [Rocznik Statystyczny Rolnictwa 2010]. W modelowym gospodarstwie uwzględniono rzeczywiste dawki nawozów mineralnych i chemicznych środków ochrony roślin, które były niższe od zalecanych norm a więc nie obniżały jakości produktów rolnych.

Wartości dochodu rolniczego i produkcji rolniczej w rozwiązaniu optymalnym modelu wielokryterialnego są o około 4% niższe od uzyskanych

w modelach jednokryterialnych. Średnio roczny dochód w gospodarstwie wraz z dopłatami wynosi 40412,51zł (161650,02/4). Przeciętne miesięczne wynagrodzenie netto (to miara porównywalna z dochodem rolniczym) w sektorze przedsiębiorstw wynosiło w 2011 r. około 2700zł [Zgierska 2011]. Uzyskany dochód rolniczy pozwala więc na godziwy poziom życia rolnika. Najbardziej opłacalne kierunki produkcji roślinnej to uprawa pszenicy i buraków cukrowych. Ze względu na duże wymagania glebowe, ich powierzchnia w modelu została ograniczona ale w rozwiązaniu, w każdym z analizowanych lat, osiągnęła górną granicę założonego udziału w strukturze zasiewów. Zarówno hodowla bydła jak i trzody chlewnej okazała się produkcją dochodową. W optymalizacji produkcji zwierzęcej nie skorzystano z programowania całkowitoliczbowego, aby uwzględnić zamknięty obrót stada zwierząt (czas użytkowania krowy: 8lat, lochy 5lat). Ułamkowe ilości sztuk zwierząt informują jaką część roku dane zwierzę przebywało w gospodarstwie oraz pozwalają uwzględnić w każdym z badanych lat koszty związane z reprodukcją stada.

Ilość substancji organicznych w glebie zmniejszyła się w rozwiązaniu optymalnym modelu wielokryterialnego czterokrotnie, mimo to jej dodatnia wielkość świadczy o nie degradacji środowiska naturalnego.

Uzyskane rozwiązanie spełnia założone oczekiwania. Istnieje możliwość uzyskania innych rozwiązań poprzez obniżanie lub podwyższanie otrzymanego dochodu lub produkcji rolniczej. Należy jednak pamiętać, aby nie nastąpiło zubożenie gleby.

WNIOSKI

1. Dochód rolniczy uzyskany w rozwiązaniu optymalnym liniowo-dynamicznego modelu wielokryterialnego jest porównywalny z wynagrodzeniem netto w sektorze przedsiębiorstw.
2. Stosowanie zasad zrównoważonego rozwoju w rolnictwie sprzyja wytwarzaniu produkcji o wysokiej jakości.
3. Zachowując równowagę między celem ekonomicznym, produkcyjnym i ekologicznym można tak zaplanować strukturę produkcji rolnej aby uzyskać dodatni bilans substancji organicznych w glebie.
4. Liniowo-dynamiczny model optymalizacji wielokryterialnej może służyć jako narzędzie wspomagające badanie zrównoważonego rozwoju na poziomie gospodarstwa rolnego.

BIBLIOGRAFIA

Augustyńska-Grzymek I. (red.) (2012) Produkcja, koszty i dochody z wybranych produktów rolniczych, Wydawnictwo Instytutu Ekonomiki Rolnictwa i Gospodarki Żywnościowej – Państwowego Instytutu Badawczego.

- Charnes A., Cooper W. (1961) *Management Models and Industrial Applications of Lineal Programming*, Wiley, Nowy Jork.
- Fotyma M., Mercik S. (1992) *Chemia rolna*, PWN, Warszawa 1992.
- Kalkulacje rolnicze (2008, 2009, 2010, 2011), Zachodniopomorski Ośrodek Doradztwa Rolniczego (biuletyny), Barzkowice .
- Krawiec B. (1991) *Metody optymalizacji w rolnictwie*, PWN, Łódź.
- Kuś J., Krasowicz S. (2001) *Przyrodniczo-organizacyjne uwarunkowania zrównoważonego rozwoju gospodarstw rolnych*, „Pam. Puł.”, z.124, Puławy.
- Rocznik Statystyczny Rolnictwa 2010
http://www.stat.gov.pl/cps/rde/xbcr/gus/rs_rocznik_rolnictwa_2010.pdf
- Staniak M. 2009, *Zrównoważony rozwój obszarów wiejskich w aspekcie środowiskowym, Woda-Środowisko-Obszary Wiejskie*, t.9,z.3(27), s 187-194.
- Szapiro T. 2001 *Decyzje menedżerskie z Excelem*, PWN.
- Zgierska A. (kierujący opracowaniem) (2011) *Przeciętne zatrudnienie i wynagrodzenie w gospodarce narodowej w 2011 r.*
http://www.stat.gov.pl/cps/rde/xbcr/gus/PW_zatrudnienie_i_wynagrodzenia_2011.pdf

MULTICRITERIAL, OPTIMIZATION MODEL FOR AN AVERAGE FARM IN ZACHODNIOPOMORSKIE PROVINCE

Abstract: Multicriterial, optimization model was created on the basis of statistical data. Its goal functions should both maximize farm income and agricultural production as well as minimize the loss of organic matter in a soil. To solve the model was used purposeful optimization. The results indicate the structure of production which (under the circumstances):

- has the highest farm income,
- provides the production of a high standard, for own use and for sale,
- prevents the degradation of the environmental long term.

The aim of the article is to study the sustainable agriculture development on the example of the average farm in Zachodniopomorskie Province. The essence of sustainable agriculture development is to maintain a balance between production, economic and environmental targets.

Keywords: multicriterial, optimization model, farm income, agricultural production

CZYNNIKI RÓŻNICUJĄCE SPOŻYCIE WYROBÓW SPIRYTUSOWYCH W POLSKICH GOSPODARSTWACH DOMOWYCH W LATACH 1999 - 2008

Krzysztof Zmarzłowski

Katedra Informatyki

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

e-mail: krzysztof_zmarzowski@sggw.pl

Grzegorz Koszela

Katedra Ekonometrii i Statystyki

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

e-mail: grzegorz_koszela@sggw.pl

Streszczenie: Celem niniejszej pracy było wyodrębnienie czynników różnicujących spożycie wyrobów spirytusowych w polskich gospodarstwach domowych w latach 1999-2008. W pracy użyto metod gradacyjnej analizy danych oraz klasycznych metod badania siły zależności zjawisk.

Słowa kluczowe: konsumpcja wyrobów spirytusowych, gradacyjna analiza danych, polskie gospodarstwa domowe, współczynnik Pearsona, współczynnik Spearmana

WSTĘP

Prawdą jest, że alkohol wielu ludziom zaszkodził, ale jakoś nikomu nie przyszło do głowy, że szkodliwe nie było używanie tego, co jest złe, ale nadużywanie tego, co bardzo dobre [Wasilewski 2007]¹. Sławne słowa Abrahama Lincolna, wypowiedziane na spotkaniu ruchu na rzecz abstynencji w Illinois w 1842 roku, odzwierciedlają kontrowersyjność tematu spożycia wyrobów alkoholowych. Konsumpcja wyrobów alkoholowych, próby jej ograniczania lub zakazu są obecne w rozwoju cywilizacji od zarania dziejów. W wielu kulturach, również w polskiej, spożywanie alkoholu bywa społecznie akceptowane i ogólnie postrzegane jako czynnik sprzyjający zacieśnianiu kontaktów międzyludzkich.

¹ Wasilewski P. *Pożytki z używki*, Polityka, Warszawa, 10 stycznia 2007

Alkohole pełnią istotną funkcję w medycynie oraz w wielu innych dziedzinach naszego życia [Woronowicz 2001]². Napoje alkoholowe odgrywają także znaczącą rolę ekonomiczną i polityczną. Spowodowane jest to dużymi przychodami tej branży i podatkami uzyskiwanymi z ich sprzedaży.

CELE, ŹRÓDŁA DANYCH I METODY BADAWCZE

Głównym celem przeprowadzonych badań było wyodrębnienie czynników ekonomiczno-społecznych, które najbardziej wpływają na zróżnicowanie spożycia wyrobów spirytusowych w polskich gospodarstwach domowych. Materiałem empirycznym wykorzystanym w badaniu była baza Głównego Urzędu Statystycznego (GUS). Dotyczyła ona budżetów polskich gospodarstw domowych (BBGD). Ze względu na brak dostępu do danych z nowszych lat w badaniach został przyjęty okres od 1999 do 2008 roku. Liczebności gospodarstw domowych w poszczególnych latach oscylowały w granicach od 31 do 37 tys. gospodarstwa domowych i była to próba reprezentatywna na skalę całej Polski. Dane, które zostały wykorzystane w badaniu były danymi miesięcznymi, a więc próba badawcza miała charakter przekrojowo-czasowy.

W badaniu dobrano arbitralnie zestaw potencjalnych czynników mogących wpływać na wielkość spożycia wyrobów spirytusowych w gospodarstwach domowych. Były to m.in.:

- Klasa miejscowości (Klasa_1: 500 tys. mieszkańców i więcej, Klasa_2: 200 – 500 tys. mieszkańców, Klasa_3: 100 – 200 tys. mieszkańców, Klasa_4: 20 – 100 tys. mieszkańców, Klasa_5: poniżej 20 tys. mieszkańców, Klasa_6: mieszkańcy wsi)
- Województwo
- Typ rodziny biologicznej (1 – małżeństwo bez dzieci, 2 – małżeństwo z 1 dzieckiem na utrzymaniu, 3 – małżeństwo z większą liczbą dzieci na utrzymaniu, 6 – matka z dziećmi na utrzymaniu, 7 – ojciec z dziećmi na utrzymaniu, 8 – inne, 9 – jednoosobowe)
- Użytkowanie gruntu (1 – tak, 2 – nie)
- Charakter zamieszkania (1 – własność dom mieszkalny lub mieszkanie, 2 – mieszkanie spółdzielcze, 3 – najem lub podnajem)
- Grupa społeczno-ekonomiczna (1 – pracowników na stanowiskach robotniczych, 2 – pracowników na stanowiskach nierobotniczych, 3 – rolników, 4 – pracujących na własny rachunek, 5 – emerytów i rencistów, 6 – utrzymujących się z niezarobkowych źródeł)
- Liczba osób w gospodarstwie
- Subiektywna ocena sytuacji materialnej (1 – bardzo dobra, 2 – raczej dobra, 3 – przeciętna, 4 – raczej zła, 5 – zła)

² Woronowicz B.T., Bez tajemnic o uzależnieniach i ich leczeniu, Instytut Psychiatrii Neurologii, Warszawa 2001

- Przychody ogółem w gospodarstwie domowym
- Rozchody ogółem w gospodarstwie domowym
- Rozchody ogółem na żywność i napoje bezalkoholowe w gospodarstwie domowym
- Spożycie ilościowe papierosów
- Płeć głowy rodziny
- Wiek głowy rodziny
- Wykształcenie głowy rodziny
- Orzeczenie inwalidztwa głowy rodziny
- Praca w pełnym wymiarze głowy rodziny (0 – nie dotyczy, 1 – pełny wymiar pracy, 2 – niepełny wymiar pracy)
- Spożycie ilościowe piwa
- Spożycie ilościowe wina

Dodatkowo zostały stworzone dwie sztuczne zmienne agregatowe dotyczące wyposażenia gospodarstwa domowego w dostępy do mediów oraz wyposażenie domostwa m.in. w sprzęt RTV, AGD, samochód i działkę rekreacyjną:

- Współczynnik zaplecza komunalnego
- Wskaźnik wyposażenia.

Do badania zależności spożycia wyrobów alkoholowych oraz czynników na nie wpływających w gospodarstwach domowych zostały wykorzystane m.in. współczynniki korelacji rang (R, Rho ρ) Spearmana oraz korelacji liniowej Pearsona. W pracy korzystano głównie ze współczynnika R Spearmana (w tym także przy znajdowaniu zależności za pomocą metod gradacyjnych), natomiast współczynnik korelacji Pearsona był używany jako uzupełnienie obliczeń. Obydwa współczynniki zostały szeroko opisane w literaturze przedmiotu^{3,4,5}, dlatego dokładniej opisano metody gradacyjnej analizy danych.

W celu wizualizacji wybranych danych dotyczących województw napisano specjalny program tworzący mapy Polski. Do gradacji kolorów na mapach zastosowano podział na 4 kategorie według popularnej metody opisanej w literaturze^{6,7}. Oparta jest ona na kryteriach zaprezentowanych w pracy wzorach (wzory 1 – 4 zawarte w Tabeli 1). Dodatkowo na niektórych mapach dołożono

³A. Luszniwicz, T. Słaby, Statystyka stosowana, PWE, Warszawa, 1996

⁴J. Józwiak, J. Podgórski, Statystyka od podstaw, PWE, Warszawa, 1992

⁵H. Kassyk-Rokicka, Statystyka nie jest trudna, PWE, Warszawa, 1992

⁶E. Nowak, Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych, PWE, Warszawa, 1990

⁷K. Zmarzłowski i inni, Porównanie ocen zróżnicowania szkolnictwa wyższego w latach 1999 – 2006 w Polsce na podstawie różnych metod normalizacji danych, Zeszyty Naukowe SGGW - Polityki Europejskie, Finanse i Marketing, Wydawnictwo SGGW, Warszawa, 2009, str.128-129

słupki procentowe, które także były odpowiednio przeskalowane do udziału procentowego danej cechy.

Tabela 1. Wzory na podstawie, których zostało sklasyfikowane spożycie ilościowe wyrobów spirytusowych w województwach

Grupa	Dolna granica	Górna granica (zamknięta z góry)	Numer wzoru
I	$\min Q_{ij}$	$\bar{Q}_{ij} - S(Q_{ij})$	(1)
II	$\bar{Q}_{ij} - S(Q_{ij})$	\bar{Q}_{ij}	(2)
III	\bar{Q}_{ij}	$\bar{Q}_{ij} + S(Q_{ij})$	(3)
IV	$\bar{Q}_{ij} + S(Q_{ij})$	$\max Q_{ij}$	(4)

gdzie: $\min Q_{ij}$ – wartość minimalna zmiennej Q_{ij} w roku,
 $\max Q_{ij}$ – wartość maksymalna zmiennej Q w roku,
 \bar{Q}_{ij} – średnia arytmetyczna wartości zmiennej Q_{ij} w roku,
 $S(Q_{ij})$ – odchylenie standardowe zmiennej Q_{ij} w roku.

W celu wykrycia zależności pomiędzy spożyciem ilościowym alkoholi a czynnikami, które wywierają na nie znaczący wpływ, wykorzystano gradacyjną analizę odpowiedności. Ogólny algorytm znajdowania zależności pomiędzy danymi można zdefiniować w kilku punktach:

1. Przygotowanie macierzy danych. W tym przypadku utworzono tablicę czynników, które zostały wyodrębnione arbitralnie. Dane te zostały podzielone osobno na każdy rok danego badania, czyli od 1999 do 2008 roku tworząc macierz.
2. Normalizacja danych. W pracy dane zostały znormalizowane w taki sposób aby suma wartości z całej macierzy elementów była równa 1.
3. Przeprowadzenie gradacyjnej analizy odpowiedności, zwanej także analizą korespondencji (GCA). Zadaniem analizy odpowiedności jest takie odtworzenie wartości liczbowych wierszy i kolumn, aby współczynnik korelacji liniowej Pearsona pomiędzy badanymi zmiennymi przyjmował tzw. lokalne maksimum (był największy). W gradacyjnej analizie odpowiedności zamiast współczynnika korelacji Pearsona brany jest pod uwagę współczynnik R Spearmana. GCA jest algorytmem iteracyjnym (powtarzany wielokrotnie)⁸. W wyniku kolejnych powtórzeń wiersze i kolumny są sortowane niemalejąco, zgodnie z otrzymanymi wartościami funkcji regresji (odnośnik). Dla ustawionych kolumn i wierszy liczony jest współczynnik R Spearmana. Jeżeli

⁸ A. Ciok, T. Kowalczyk, E. Pleszczyńska, W. Szczesny. Algorithms of grade correspondence–cluster analysis. The Collected Papers of Theoretical and Applied Computer Science. Vol. 6, Nr. 1-4, 5-22, 1995

- wartości tego współczynnika w kolejnych powtórzeniach zaczynają zmierzać do tej samej wartości, oznacza to, że procedura znalazła optymalne ustawienie wartości w kolumnach i wierszach względem siebie [Kowalczyk i in. 2004],
4. Przeprowadzenie analizy skupień oraz agregacji danych. Analiza skupień jest zbiorem metod stosowanych w celu utworzenia grup wartości lub cech, które są do siebie podobne i tworzą tzw. skupienia. Gradacyjna analiza skupień oparta jest na dobraniu przez badacza takiej liczby skupień, która zmaksymalizuje wskaźnik R Spearmana. Dokładny algorytm matematyczny tej metody jest opisany w publikacjach z tego zakresu [Matyja i in. 2005].
 5. Wygenerowanie wyników liczbowych oraz map nadreprezentacji. Mapa nadreprezentacji jest graficznym obrazem przeprowadzonej procedury. W najprostszym ujęciu w kolumnach posortowane są zmienne w sposób biegunowy, tzn. najbardziej na lewo (lub prawo), w zależności od ustawienia badacza zwykle umieszczona zostaje zmienna zależna (w tym przypadku spożycie wyrobów spirytusowych w gospodarstwach domowych), a najbardziej na prawo zmienna, która była najbardziej ujemnie skorelowana z danym zjawiskiem. Dodatkowo szerokość każdej z kolumn może powiedzieć nam o istotnym znaczeniu danej cechy w całym zbiorze danych (im szersza kolumna, tym większy udział zmiennej w całej macierzy danych). Komórki danej mapy są pokolorowane tak, aby również pokazywać nadreprezentacje pewnych obserwacji względem teoretycznego rozkładu. Nadreprezentacja pokazuje, które wartości zmiennej wyjaśniają badane zjawisko w najlepszym stopniu (im ciemniejsze komórki, tym obserwacje lub ich skupienia mają większy wpływ na kształtowanie się danego zjawiska), czyli których tak naprawdę jest najwięcej w zbiorze,
 6. Wyznaczenie macierzy korelacji rangowych R Spearmana w celu potwierdzenia rozpoznanych zależności [Kowalczyk i in. 2004].

WYNIKI BADAŃ

W celu wyodrębnienia zmiennych wpływających na spożycie wyrobów spirytusowych w gospodarstwach domowych, posłużono się gradacyjną analizą odpowiedności oraz współczynnikami miar korelacji Pearsona i rangowej Spearmana⁹. Ponieważ wyniki analizy korelacji dla poszczególnych lat były zbliżone do siebie, dlatego w tabelach zamieszczono jedynie wyniki dla pierwszego i ostatniego badanego roku. Do metod gradacyjnych został użyty program statystyczny Gradestat. Ze względu na możliwości obliczeniowe programu Gradestat wartości niektórych zmiennych musiały zostać wyrażone w kwartylach i w kwintylach. Były to m.in.: spożycie ilościowe wyrobów spirytusowych, piwa, wina, papierosów (kwintyle), przychody ogółem, rozchody

⁹ Miary korelacji zostały zastosowane w przypadkach zmiennych ilościowych oraz porządkowych.

ogółem, rozchody na artykuły żywnościowe i napoje bezalkoholowe, wiek głowy rodziny, wskaźniki wyposażenia oraz wyposażenia komunalnego (kwartyle). W przypadku spożycia ilościowego poszczególnych alkoholi i papierosów wyodrębniono kwintyle, w tym sztuczną grupę kwintylową (kwintyl 0)¹⁰. Znalazły się w niej gospodarstwa deklarujące zerowe spożycie danego wyrobu alkoholowego. Współczynniki korelacji Pearsona i Spearmana wyliczono na podstawie surowych danych, tzn. bez podziału na kwartale i kwintyle. Ponieważ niektóre zmienne miały charakter jakościowy użycie analizy korelacji w ich przypadku nie miało żadnego sensu. W takich przypadkach przeprowadzono jedynie gradacyjną analizę odpowiedności. Wyniki tej analizy zostały podzielone na 5 skupień. Pierwsze dwa zawierały czynniki mające największy wpływ na ilościową konsumpcję danego alkoholu.

Tabela 2. Wyniki gradacyjnej analizy danych oraz współczynników korelacji Pearsona oraz Spearmana dla lat 1999 i 2008 – wyroby spirytusowe

Zmienna	1999		2008	
	Korelacja liniowa Pearsona	Korelacja rangowa Spearmana	Korelacja liniowa Pearsona	Korelacja rangowa Spearmana
Przychody ogółem w gospodarstwie domowym	0,15	0,24	0,12	0,24
Rozchody ogółem w gospodarstwie domowym	0,18	0,26	0,16	0,25
Rozchody ogółem na żywność i napoje bezalkoholowe w gospodarstwie domowym	0,17	0,20	0,16	0,20
Wskaźnik wyposażenia	0,12	0,15	0,11	0,17
Wskaźnik wyposażenia komunalnego	0,07	0,08	0,05	0,09
Spożycie ilościowe piwa	0,24	0,33	0,24	0,32
Spożycie ilościowe wina	0,21	0,27	0,22	0,23
Subiektywna ocena sytuacji materialnej	-0,11	-0,15	-0,09	-0,14
Spożycie ilościowe papierosów	0,09	0,12	0,07	0,12
Płeć głowy rodziny	-0,08	-0,12	-0,07	-0,11
Wykształcenie głowy rodziny	-0,05	-0,08	-0,05	-0,09
Typ rodziny biologicznej	-	-	-	-
Grupa społeczno-ekonomiczna	-	-	-	-

Źródło: opracowanie własne na podstawie danych GUS

¹⁰ Ze względu na stworzenie 0 kwintyla numerację poszczególnych kwintyli rozpoczęto od 0.

Kryterium wyboru czynników powiązanych ze spożyciem ilościowym alkoholi określono za pomocą liczby wystąpień danego typu czynnika w badanym okresie. Granicę ustalono na poziomie 5 z 10 badanych lat, czyli np.: jeżeli 4 kwartył przychodów ogółem w 5 lub więcej latach odznaczał się zależnością ze spożyciem ilościowym danego alkoholu, klasyfikowano go jako czynnik mający wpływ na ilość konsumowanego alkoholu. W przeciwnym przypadku nie był traktowany jako pojedyncze wystąpienie i był pomijany w dalszej analizie.

W tabeli 2 zamieszczono czynniki wyodrębnione za pomocą gradacyjnej analizy odpowiedności, jako te, które wpływają na większe spożycie wyrobów spirytusowych w gospodarstwach domowych.

Uzyskane wyniki wskazują na to, że ilościową konsumpcję wyrobów spirytusowych najbardziej preferują zamożniejsze domostwa. Przez cały badany okres gospodarstwa, które uzyskiwały największe przychody brutto, częściej spożywały tego rodzaju produkty. 3 i 4 kwartył przychodów ogółem w pierwszym i drugim skupieniu znalazł się w wszystkich badanych latach. Potwierdzają to również dodatnie i istotnie statystycznie współczynniki korelacji Pearsona oraz Spearmana. Znaczenie zamożności gospodarstwa domowego względem ilości wypijanych wyrobów spirytusowych potwierdzają także:

- subiektywna ocena sytuacji materialnej, która im lepsza tym więcej spirytusów wypijano. Potwierdzają to ujemne wartości współczynników korelacji Pearsona i Spearmana¹¹. W pierwszym i drugim skupieniu zmiennych, zmienna określająca, że sytuacja materialna gospodarstwa była bardzo dobra wystąpiła we wszystkich badanych latach, a raczej dobra w 8 badanych latach,
- wskaźniki wyposażenia i wyposażenia komunalnego. Dla tych dwóch zmiennych wyniki analizy gradacyjnej i statystycznie istotne dodatnie współczynniki korelacji oznaczają, że wraz ze wzrostem wartości tych współczynników rośnie spożycie ilościowe wyrobów spirytusowych. Również gradacyjna analiza odpowiedności wykazała, że 3 i 4 kwartył wartości tych współczynników znalazły się odpowiednio przez 9 i 5 lat w pierwszych dwóch skupieniach zmiennych mających największy wpływ na spożycie ilościowe spirytusów w gospodarstwach domowych.

Kolejnymi zmiennymi znacząco wpływającymi na ilość konsumowanych wyrobów spirytusowych były rozchody finansowe ogółem i rozchody finansowe na artykuły żywnościowe w domostwach. Potwierdzają zarówno wyniki analizy gradacyjnej (we wszystkich badanych latach w pierwszych dwóch skupieniach znalazły się gospodarstwa z 3 i 4 kwartyła rozchodów ogółem i 4 kwartyła rozchodów na artykuły żywnościowe), jak i analiza korelacji. Współczynniki skorelowania pomiędzy ilościową konsumpcją spirytusów, a wielkością wydatków

¹¹ Należy zwrócić uwagę, że ocena sytuacji materialnej była uporządkowana w sposób malejący, czyli 1 - najlepsza, a 5- najgorsza, dlatego współczynniki skorelowania miały ujemne znaki.

były we wszystkich latach dodatnie. Oznacza to, iż im większe wydatki ogółem oraz na żywność, tym częściej też wydajemy na alkohole spirytusowe.

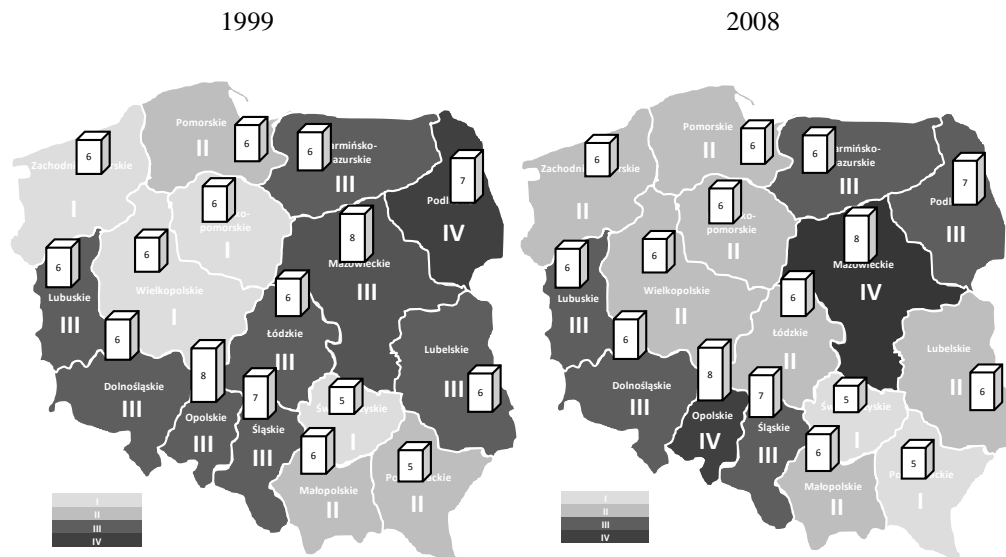
Związek zaplecza materialnego z ilością konsumowanych spirytusów widoczny jest w poziomie posiadanego wykształcenia głowy rodziny. Zdecydowanie najwięcej wyrobów spirytusowych konsumowano w gospodarstwach, w których głowa rodziny miała wykształcenie wyższe. Takie wyniki mogą wskazywać na związek pomiędzy wykształceniem głowy rodziny oraz uzyskiwanymi przychodami gospodarstw domowych. We wszystkich latach wykształcenie wyższe znajdował się w 1 i 2 skupieniu zmiennych, a przez 8 lat również wykształcenie średnie zawodowe. Najmniej spożywały gospodarstwa domowe, w których głowa gospodarstwa domowego nie ma żadnego wykształcenia lub podstawowe czy też gimnazjalne. Wynik ten może także sugerować, że gospodarstwa z bardziej wykształconymi członkami rodziny częściej przyznają się do kupowania oraz picia alkoholu, niż te o niższym poziomie wykształcenia.

Gospodarstwa domowe spożywające większe ilości alkoholi spirytusowych, preferowały także konsumpcję większych ilości piwa oraz wina. Dla piwa analiza gradacji wykazała, że przez 10 lat domostwa z 3 i 4 kwintyla ilościowego spożycia piwa konsumowały również najwięcej wyrobów spirytusowych. W przypadku wina analiza gradacyjna nie dała jednoznacznej odpowiedzi, czy gospodarstwa pijące więcej wina piją więcej spirytusów (wszystkie kwintyle spożycia wina znalazły się w pierwszych dwóch skupieniach). W tym celu dodatkowy kryterium oceny były wartości współczynników korelacji, które wykazały dodatnią zależność pomiędzy spożyciem ilościowym wyrobów spirytusowych zarówno wina jak i piwa.

Palenie papierosów, również okazało się dodatnio skorelowane z ilością spożywanych spirytusów. Według gradacyjnej analizy odpowiedności przez 7 lat 4 kwintyl oraz przez 10 lat 3 kwintyl konsumpcji papierosów. Prowadzi to do stwierdzenia, że osoby, które częściej piją alkohole spirytusowe lubią sięgać po papierosy. Dodatnią zależność pomiędzy paleniem i piciem obrazują także dodatnie współczynniki korelacji, które dla tych dwóch czynników był istotne statystycznie we wszystkich badanych latach.

Biorąc pod uwagę czynniki demograficzne wyraźnie wyróżniła się płeć głowy gospodarstwa. Okazuje się, że tam, gdzie głową rodziny była mężczyzna pito większe ilości spirytusów. W przypadku natomiast typu biologicznego rodziny zdecydowanie częściej sięgały po spirytusy małżeństwa bez dzieci lub posiadające wyłącznie jedno dziecko. Obydwie te zmienne przez 10 badanych lat znajdowały się w pierwszych dwóch skupieniach. Niestety w tym przypadku policzenie współczynników korelacji nie miało najmniejszego sensu ze względu na brak możliwości interpretacji wyników. W przypadku grupy społeczno-ekonomicznej zdecydowanymi liderami były gospodarstwa rolników (10 lat w 1 i 2 skupieniu), pracowników na stanowiskach nierobotniczych oraz pracujących na własny rachunek.

Rysunek 1. Klasy spożycia wyrobów spirytusowych w 1999 i 2008 roku łącznie z procentowym rozkładem spożycia w poszczególnych województwach



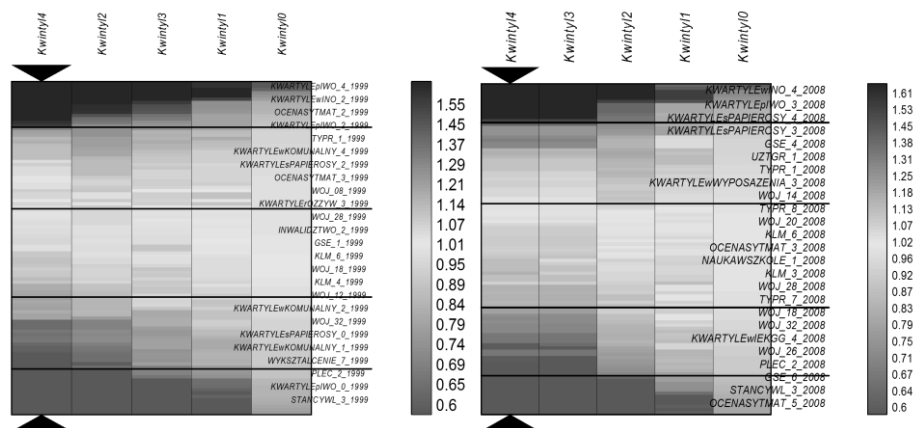
Źródło: Opracowanie własne na podstawie danych GUS

W przypadku województw na podstawie analizy gradacyjnej wyróżniono w szczególności jedno województwo - podlaskie. Analizując 1999 rok, można zauważyć, że właśnie w tym województwie wystąpiło największe spożycie alkoholi spirytusowych (IV najwyższa klasa spożycia ilościowego oraz 9% udział całkowitego spożycia gospodarstw) (patrz: Rysunek 1). W III klasie ilościowej konsumpcji spirytusów znalazły się takie województwa jak: dolnośląskie, lubelskie, lubuskie, łódzkie, mazowieckie, opolskie, śląskie, warmińsko-mazurskie. Najmniejsze spożycie odnotowano w gospodarstwach domowych z województw: kujawsko-pomorskiego, świętokrzyskiego, wielkopolskiego oraz zachodniopomorskiego. W roku 2008 sytuacja uległa zmianie, gdyż zdecydowanie najwięcej wypijano w domostwach z województw: mazowieckiego i opolskiego (patrz: Rysunek 1). W III klasie spożycia ilościowego znalazły się województwa: dolnośląskie, lubuskie, podlaskie, śląskie oraz warmińsko-mazurskie. Najniższy poziom konsumpcji spirytusów zadeklarowały gospodarstwa domowe w województwach: podkarpackim oraz świętokrzyskim. Analiza korelacji oraz analiza gradacyjna nie wyróżniły klasy miejsca zamieszkania jako czynnika mającego duże znaczenie do ilości konsumowanych alkoholi spirytusowych.

Mapy nadreprezentacji są wizualizacją wyników i zaprezentowano je wyłącznie dla 1 i ostatniego roku badania. Mapy nadreprezentacji dla kwintyli spożycia wyrobów spirytusowych względem badanych zmiennych dla lat 1999 i 2008 prezentuje rysunek **Błąd! Nie można odnaleźć źródła odwołania..** Na órnej krawędzi mapy znajdują się kwintyle spożycia ilościowego alkoholi

spirytusowych w gospodarstwach domowych. Na prawej bocznej krawędzi znalazły się etykiety zmiennych (ze względów, które były brane pod uwagę w badaniu (ze względu na dużą ilość przyjmowanych wariantów aplikacja do tworzenia map nie była w stanie umieścić wszystkich etykiet zmiennych)). Poziome linie mapy oznaczają granice skupień na jakie zostały te zmienne podzielone. Kolorystyka mapy z kolei mówi nam o tym z jaką siłą dana zmienna wyjaśnia spożycie spirytusów w poszczególnych kwintylach, czyli jest tak zwaną nadreprezentacją w sytuacji silnej zależności dodatniej oraz niedoreprezentacją w odwrotnym przypadku. Na mapach nadreprezentacji czynniki, które są ustawione po przekątnej od lewego górnego rogu (w tym przypadku jest to 4 kwintyl spożycia ilościowego wyrobów spirytusowych) do prawego dolnego są w najmniejszym stopniu od siebie zależne. Natomiast te bliżej prawej górnej strony są najbardziej powiązane ze najwyższym kwintylem spożycia ilościowego spirytusów.

Rysunek 2. Mapy nadreprezentacji czynników wpływających ilościowe spożycie wyrobów spirytusowych (podzielonego na kwintyle) dla lat 1999 i 2008¹²



Źródło: Opracowanie własne na podstawie danych GUS

WNIOSKI

1. Zdecydowanie najbardziej wpływającym czynnikiem na ilość wypijanych wyrobów spirytusowych w gospodarstwach domowych jest zaplecze finansowe domostw. Rodziny mające większe przychody, lepiej wyposażone oraz pozytywnie oceniające swoją sytuację materialną konsumowały więcej tego rodzaju alkoholu.

¹² Poziome linie obrazują podział na 5 skupień zmiennych. Pierwsze dwie linie od góry są granicą dwóch skupień, w których znalazły się zmienne najbardziej skorelowane z najwyższym kwintylem spożycia ilościowego wyrobów spirytusowych.

2. Gospodarstwa domowe, które deklarowały spożycie wyrobów spirytusowych, równie chętnie sięgały po piwo i wino.
3. Domostwa z „męską głową rodziny” zdecydowanie preferują spożycie wyrobów spirytusowych. Domostwa, które deklarowały spożywanie wyrobów spirytusowych często sięgały również po papierosy. Zależność ta odznaczała się dodatnią korelacją pomiędzy tymi spożyciem ilościowym tych alkoholi i konsumowanych papierosów.
4. Najwięcej wyrobów spirytusowych wypijano w gospodarstwach, w których głowa rodziny miała wykształcenie wyższe, natomiast najmniej w tych gdzie głowa rodziny była najslabiej wykształcona. Z powyższego wynika, że na wysokość uzyskiwanych dochodów w dużej mierze wpływ ma poziom wykształcenia danej osoby.
5. Dodatkowo wyroby spirytusowe były dość popularne wśród gospodarstw domowych rolników. Alkohole te najrzadziej pite były w gospodarstwach emerytów i rencistów oraz osób utrzymujących się z niezarobkowych źródeł.
6. Wyroby spirytusowe najczęściej są używką w domostwach z większych aglomeracji miejskich oraz wsi. Najmniej spirytusów wypito w gospodarstwach z niedużych miast.
7. Ze względu na zróżnicowanie geograficzne najczęściej wyrobów spirytusowych wypijały gospodarstwa domowe z województw: podlaskiego, mazowieckiego i lubelskiego. Najmniej z województw: kujawsko-pomorskiego, świętokrzyskiego i wielkopolskiego.

BIBLIOGRAFIA

- Ciok A., Kowalczyk T., Pleszczyńska E., Szczesny W. (1995) Algorithms of grade correspondence-cluster analysis. „The Collected Papers of Theoretical and Applied Computer Science”. Vol. 6, Nr. 1-4, 5-22.
- Józwiak J., Podgórski J., (1992) Statystyka od podstaw, PWE, Warszawa.
- Kassyk-Rokicka H. (1992) Statystyka nie jest trudna, PWE, Warszawa.
- Kowalczyk T., Pleszczyńska E., Ruland F. (2004) Grade Models and Methods for Data Analysis: With Applications for the Analysis of Data Populations, Studies in Fuzziness and Soft Computing, vol. 151, Springer-Verlag, str. 271-274.
- Luszniewicz A., Słaby T. (1996) Statystyka stosowana, PWE, Warszawa.
- Matyja O., Pleszczyńska E., Wiech M. (2005) Analiza danych medycznych i demograficznych przy użyciu programu GradeStat, Instytut Podstaw Informatyki PAN, Instytut "Pomnik - Centrum Zdrowia Dziecka", Warszawa, str. 21-23.
- Nowak E. (1990) Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych, PWE, Warszawa.
- Wasilewski P. (2007) Pożytki z używki, Polityka, Warszawa, 10 stycznia 2007.
- Woronowicz B.T. (2001) Bez tajemnic o uzależnieniach i ich leczeniu, Instytut Psychiatrii Neurologii, Warszawa.

Zmarzłowski K. i inni (2009) Porównanie ocen zróżnicowania szkolnictwa wyższego w latach 1999 – 2006 w Polsce na podstawie różnych metod normalizacji danych, Zeszyty Naukowe SGGW - Polityki Europejskie, Finanse i Marketing, Wydawnictwo SGGW, Warszawa, str.128-129.

FACTORS INFLUENCING ALCOHOL CONSUMPTION IN POLISH HOUSEHOLDS IN 1999-2008

Abstract: The main objective of this research was to determine factors that influenced consumption of alcoholic beverage in Polish households in 1999-2008. During the course of study authors applied classic methods of correlation and modern methods of grade data analysis.

Keywords: Alcohol consumption, Grade data analysis, Polish households, Pearson correlation coefficient, Spearman rank correlation coefficient