

METODY ILOŚCIOWE  
W BADANIACH EKONOMICZNYCH

QUANTITATIVE METHODS  
IN ECONOMICS

Vol. XVIII, No. 4

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

**QUANTITATIVE METHODS  
IN ECONOMICS**

**METODY ILOŚCIOWE W BADANIACH  
EKONOMICZNYCH**

**Volume XVIII, No. 4**

Warsaw 2017

## **EDITORIAL BOARD**

Editor-in-Chief: Bolesław Borkowski

Vice-Editor-in-Chief: Hanna Dudek

Managing Editor: Grzegorz Koszela

Theme Editors:

Econometrics: Bolesław Borkowski

Multidimensional Data Analysis: Wiesław Szczesny

Mathematical Economy: Zbigniew Binderman

Analysis of Labour Market: Joanna Landmesser

Financial Engineering: Monika Krawiec

Data Science: Michał Gostkowski

Statistical Editor: Wojciech Zieliński

Technical Editors: Jolanta Kotlarska, Elżbieta Saganowska

Language Editor: Agata Kropiwić

Native Speaker: Yochanan Shachmurove

Editorial Assistant: Luiza Ochnio

## **SCIENTIFIC BOARD**

Adnene Ajimi (University of Sousse, Tunisia)

Heni Boubaker (University of Sousse, Tunisia)

Vasily Dikussar (Doradnicyn Computing Centre of the Russian Academy of Sciences, Russia)

Peter Friedrich (University of Tartu, Estonia)

Paolo Gajo (University of Florence, Italy)

Agnieszka Gehringer (University of Göttingen, Germany)

Anna Maria Gil-Lafuente (University of Barcelona, Spain)

Jaime Gil-Lafuente (University of Barcelona, Spain)

Vasile Glavan (Moldova State University, Moldova)

Francesca Greselin (The University of Milano-Bicocca, Italy)

Ana Kapaj (Agriculture University of Tirana, Albania)

Jirawan Kitchaicharoen (Chiang Mai University, Thailand)

Yuriy Kondratenko (Black Sea State University, Ukraine)

Vassilis Kostoglou (Alexander Technological Educational Institute of Thessaloniki, Greece)

Karol Kukula (University of Agriculture in Krakow, Poland)

Kesra Nermend (University of Szczecin, Poland)

Nikolas N. Olenev (Doradnicyn Computing Centre of the Russian Academy of Sciences, Russia)

Alexander N. Prokopenya (Brest State Technical University, Belarus)

Yochanan Shachmurove (The City College of The City University of New York, USA)

Mirbulat B. Sikhov (al-Farabi Kazakh National University, Kazakhstan)

Marina Z. Solesvik (Nord University, Norway)

Ewa Syczewska (Warsaw School of Economics, Poland)

Achille Vernizzi (University of Milan, Italy)

Andrzej Wiatrak (University of Warsaw, Poland)

Dorota Witkowska (University of Lodz, Poland)

ISSN 2082-792X

e-ISSN 2543-8565

© Copyright by Department of Econometrics and Statistics WULS – SGGW  
(Katedra Ekonometrii i Statystyki SGGW)

Warsaw 2017, Volume XVIII, No.4

The original version is the paper version

Journal homepage: [qme.sggw.pl](http://qme.sggw.pl)

Published by Warsaw University of Life Sciences Press

## SPIS TREŚCI

Marek Angowski, Katarzyna Domańska, Tomasz Kijek – Zastosowanie metody k-średnich w segmentacji nabywców na rynku produktów spożywczych .....	521
Beata Bieszk-Stolorz, Iwona Markowicz – Ocena prawdopodobieństwa rezygnacji z pośrednictwa urzędu pracy w poszukiwaniu zatrudnienia .....	531
Krzysztof Borowski – Rozkład normalny stóp zwrotu z akcji wchodzących w skład następujących indeksów giełdowych: WIG20, mWIG40 i sWIG80 .....	541
Jadwiga Bożek, Karol Kukuła, Czesław Nowak – Zróżnicowanie i dynamika zmian powierzchni użytków rolnych według grup obszarowych gospodarstw w krajach Unii Europejskiej w okresie 2010-2013 .....	561
Iwona Foryś – Aktywność zawodowa kobiet na polskim rynku nieruchomości na przykładzie zawodów regulowanych .....	573
Wiesław Grygierzec – Własności funkcji wartości dla stochastycznego problemu sterowania optymalnego typu Mayera .....	584
Alicja Grześkowiak – Analiza zasobów kompetencji z wykorzystaniem grupowania zmiennych niemetrycznych .....	592
Małgorzata Kobylińska – Zanurzanie obserwacji w próbie w ocenie zróżnicowania przestępczości przeciwko mieniu oraz stopy bezrobocia w Polsce .....	602
Krzysztof Kompa, Dorota Witkowska – Wpływ zmian frakcji kobiet w zarządzaniu na zmiany ROE spółek publicznych .....	614
Dominik Krężolek – Skośność rozkładu a estymacja kwantylowych miar ryzyka – przypadek rynku metali .....	624
Sławomir Lisek – Struktura wielkościowa przedsiębiorstw w Polsce .....	635
Jan Malchar, Monika Zielińska-Sitkiewicz – Metody klasyfikacji w analizie porównawczej rozwoju społeczno-gospodarczego polskich województw w latach 2010 i 2014 – wpływ procedury normalizacji na wynik rankingu .....	643
Iwona Markowicz, Beata Bieszk-Stolorz – Metody analizy trwania w identyfikacji determinant ryzyka wykreślenia z rejestru bezrobotnych .....	653
Aleksandra Pawłowska – Współoddziaływanie wybranych dopłat na wzrost wydajności pracy w polskich gospodarstwach rolnych .....	663
Michał Staszak – Eksperymentalna ocena efektywności portfela fundamentalnego dla spółek z indeksu WIG20 za lata 2004-2016 .....	672

---

Ryszard Szupiluk, Paweł Rubach – Identyfikacja komponentów destrukcyjnych w modelach predykcyjnych w podejściu wielomodelowym.....	679
Agnieszka Tłuczak – Dynamiczna analiza shift share – ocena zmian struktury produkcji zwierzęcej w Polsce .....	689
Krzysztof Zmarzłowski – Metody prezentacji danych w systemach BI .....	698

## ZASTOSOWANIE METODY K-ŚREDNICH W SEGMENTACJI NABYWCÓW NA RYNKU PRODUKTÓW SPOŻYWCZYCH

**Marek Angowski (ORCID: 0000-0003-1669-5073)**

**Katarzyna Domańska**

Wydział Agrobiotechnologii

Uniwersytet Przyrodniczy w Lublinie

e-mail: marek.angowski@up.lublin.pl; katarzyna.domanska@up.lublin.pl

**Tomasz Kijek (ORCID: 0000-0002-0134-4943)**

Wydział Ekonomiczny

Uniwersytet Marii Curie Skłodowskiej w Lublinie

e-mail: tomasz.kijek@poczta.umcs.lublin.pl

**Streszczenie:** Poznanie preferencji i zachowań nabywców jest kluczowym czynnikiem decydującym o sukcesie rynkowym przedsiębiorstwa. Skuteczny marketing wymaga precyzyjnej segmentacji, zgodnie z którą tworzy się spersonalizowaną ofertę rynkową uwzględniającą możliwości i potrzeby konkretnej grupy docelowej. Metoda k-średnich jest metodą badawczą, która może stanowić podstawę segmentacji nabywców, ponieważ odzwierciedla reakcje konsumentów na poszczególne warianty lub atrybuty produktów. Celem publikacji jest prezentacja możliwości aplikacji metody k-średnich w procedurze segmentacji na rynku produktów spożywczych.

**Słowa kluczowe:** segmentacja rynku, metoda k-średnich, profile nabywców

### WSTĘP

Podstawą strategii marketingowej jest zorientowanie wszystkich działań i zasobów przedsiębiorstwa na potrzeby, wymagania i preferencje aktualnych oraz potencjalnych konsumentów. Ważne jest zatem regularne pozyskiwanie i analizowanie informacji o konsumentach oraz ich zachowaniach na rynku.

Zachowania konsumentów na rynku związane są z wyborem, zakupem, użytkowaniem i oceną produktów lub usług, w rezultacie których nabywcy spodziewają się zaspokojenia odczuwanych potrzeb. Przedmiotem badań nad zachowaniem konsumentów są więc procesy podejmowania decyzji o przez-

naczeniu posiadanych zasobów – środków pieniężnych, czasu i wysiłku – na cele związane z konsumpcją. Prowadzą one do znalezienia odpowiedzi na pytania: co konsumenci kupują, dlaczego, w jaki sposób, gdzie, kiedy, z jaką częstotliwością, w jakich ilościach, z jakim przeznaczeniem, jakie są ich wymagania i oczekiwania związane z jakością produktów [Venter i in. 2015].

Potrzeby, preferencje i zachowania nabywców tworzących dany rynek z reguły wykazują pewne zróżnicowanie. Wynika z tego konieczność ustalenia sylwetki (profilu) typowego nabywcy, który stanowiłby punkt odniesienia zarówno w ocenie dotychczasowych działań marketingowych, jak i planowaniu nowych przedsięwzięć na rynku. Konieczne jest więc szukanie wspólnych cech ujawniających się w zachowaniach właściwych pewnym grupom konsumentów, co prowadzi do określenia segmentacji rynku [Lipowski i in. 2016].

Głównym celem publikacji jest przedstawienie możliwości wykorzystania metody k-średnich w procedurze segmentacji na rynku produktów spożywczych.

## ISTOTA I POJĘCIE SEGMENTACJI RYNKI

Segmentacja rynku to podział rynku według określonych kryteriów na względnie jednorodne (homogeniczne) grupy nabywców (segmenty rynku), które wyznaczają dla przedsiębiorstwa obszar działania i mogą stanowić punkt odniesienia przy kształtowaniu programu marketingowego. Grupowanie nabywców w określony sposób jest istotą segmentacji rynku. Konieczność segmentacji wynika przede wszystkim z występowania następujących przesłanek [Dębkowska 2010, Tonks 2009]:

- nie wszyscy nabywcy są tacy sami,
- z grupy nabywców określonego produktu można wyodrębnić ich podgrupy o podobnych zachowaniach, systemach wartości itp.,
- podgrupy są liczebnie mniejsze i bardziej homogeniczne niż grupa jako całość,
- łatwiejsze i skuteczniejsze jest działanie na rzecz mniejszej podgrupy podobnych konsumentów niż na rzecz dużych, zróżnicowanych grup.

Poprawnie wyodrębniony segment rynku powinien się odznaczać następującymi cechami [Garbarski i in. 2006, Stanimir 2006]:

- wymierność: można zmierzyć wielkość, siłę nabywczą i profil segmentów rynku (cechy konsumentów odróżniające ich od całego rynku),
- rozległość: powinien być tak rozległy, aby uzasadniać zastosowanie indywidualnej strategii marketingowej,
- dostępność: do segmentu można skutecznie dotrzeć i go obsłużyć,
- możliwość odróżnienia: segmenty dają się odróżnić i w odmienny sposób reagują na poszczególne elementy marketing mix i strategię marketingowe,
- możliwość działania: można stworzyć skuteczne programy obsługi segmentów.

Poprawnie przeprowadzona segmentacja rynku umożliwia przedsiębiorstwu [Maricic i in. 2015, Hunt i in. 2004, Reynolds 2006]:



- dostosowanie produktów do potrzeb nabywców, co przyczynia się do zwiększenia sprzedaży i konkurencyjności danych towarów na rynku,
- dotarcie do tych konsumentów, których preferencje są uwzględniane przy kształtowaniu produktu,
- efektywniejsze wydatkowanie środków finansowych przeznaczonych na działania marketingowe przez odpowiedni dobór instrumentów w stosunku do określonych i znanych nabywców,
- szybkie zauważanie zmian na rynku dzięki ciągłym obserwacjom określonych jego części oraz wcześniejsze przygotowanie się do tych zmian i odpowiednie na nie reagowanie.

Segmenty rynku mogą być wyodrębnione za pomocą dwóch metod. Pierwsza polega na zastosowaniu wybranych kryteriów a priori. W takim przypadku badający z góry ustala kryteria podziału rynku. Na podstawie przeprowadzonego badania ustala się potencjalną wielkość rynku oraz charakterystykę klientów, wykorzystując m. in. kryteria geograficzne, demograficzne, psychologiczne, behawioralne. Metoda a priori umożliwia tylko pokazanie charakterystyki wybranego segmentu rynku, nie można natomiast dokonać pomiaru zyskowności tego segmentu ani jego wielkości w porównaniu z innymi. Innym ograniczeniem jest to, że bazuje na zewnętrznych charakterystykach klientów, które wcale nie muszą być tymi czynnikami, które determinują postępowanie klientów [Maciejewski 2012, Walesiak 2000].

Druga metoda polega na segmentacji rynku post hoc. Opiera się ona na podziale rynku na podstawie zbioru cech określonych empirycznie w wyniku przeprowadzonych badań rynku. Wykorzystuje się tu metody statystyczne, które pozwalają na bardziej kompleksowe spojrzenie na docelowe grupy nabywców produktu, ich liczebność, wymagania i preferencje, dostępność czy wrażliwość na oddziaływanie instrumentów marketingowych [Żakowska-Biemans i in. 2013]

Wśród taksonomicznych metod segmentacji rynku, który nie ma charakteru rozłącznego można wskazać następujące przykłady: metody hierarchiczne (np. metoda Warda), metody niehierarchiczne (np. k-średnich), metody aglomeracyjne (np. metoda najbliższego sąsiedztwa), metody podziałowe (np. metoda Czekanowskiego), metody obszarowe (np. metoda Hartigana), metody optymalizacyjne (np. metoda Wisharta).

Skuteczność segmentacji rynku zależy od przyjęcia odpowiednich kryteriów, które pozwolą wyjaśnić różnice w zachowaniach między grupami nabywców. W literaturze przedmiotu spotykamy bardzo dużą liczbę szczegółowych kryteriów opisujących rynek i konsumentów, uwzględniających szersze i węższe ujęcia, bardziej ogólne i szczegółowe, odnoszących się do nabywców dóbr i usług konsumpcyjnych, do nabywców dóbr i usług przemysłowych [Mazurek-Łopacińska 2005]. Najczęściej spotykana klasyfikacja kryteriów segmentacji rynku odnośnie konsumentów indywidualnych dzieli je na kryteria odnoszące się do lokalizacji (geograficzne), konsumenta (demograficzne, społeczno-ekonomiczne,

psychologiczne) oraz kryteria odnoszące się do produktu lub sytuacji zakupu (behawioralne) [Han i in. 2014, Mahajan i in. 2008, Verain 2012].

Niezależnie od typu oraz charakteru zastosowanych kryteriów należy pamiętać o dwóch podstawowych zasadach dokonywania segmentacji. Po pierwsze, nie ma kryteriów uniwersalnych, nadających się do zastosowania w odniesieniu do każdego produktu i każdej sytuacji rynkowej. Po drugie, dokonywanie segmentacji rynku jest procesem ciągłym, wymagającym stałego badania cech konsumentów i różnic pomiędzy nimi oraz dostosowywania programu działania do zachodzących zmian [Kotler 2005, Muller i in. 2014].

## WYKORZYSTANE DANE

Część empiryczna została opracowana na podstawie wyników badań sondażowych, które były przeprowadzone w 2016 roku na grupie 358 studentów przy wykorzystaniu metody doboru celowego. Główną metodą zbierania danych była metoda CAWI wspomagana metodą PAPI. Kwestionariusz został przygotowany w wersji elektronicznej i powiązany z bazą danych (SurveyMonkey). Kwestionariusz składał się z 18 pytań problemowych dotyczących oceny zachowań i postaw nabywców na rynku produktów żywnościowych ze szczególnym uwzględnieniem czynników kształtujących proces wyboru oraz pytania charakteryzujące respondenta. W analizie głównych składowych wykorzystano oryginalne dane w postaci odpowiedzi respondentów na skali porządkowej. Badania dotyczyły czterech grup produktowych, m.in. mięsa i jego przetworów.

## METODA BADAWCZA

W celu przeprowadzenia segmentacji rynku nabywców mięsa i jego przetworów zaproponowano procedurę badawczą obejmującą następujące etapy/czynności:

1. Zredukowanie liczby zmiennych charakteryzujących preferencje zakupowe nabywców za pomocą analizy głównych składowych.
2. Wyznaczenie liczby segmentów przy wykorzystaniu metody Warda z regułą stop opartą na indeksie Calińskiego–Harabasza.
3. Przeprowadzenie segmentacji rynku w oparciu o metodę k-średnich.

Metoda głównych składowych wykorzystana w pierwszym etapie procesu badawczego polega na transformacji obserwowalnych zmiennych pierwotnych w nowe wzajemnie ortogonalne zmienne, tzw. główne składowe [Gatnar i in. 2004]. W zapisie macierzowym model analizy głównych składowych przyjmuje postać układu równań:

$$\begin{aligned} Z^T &= BS^T \\ S &= B^T Z \end{aligned} \tag{1}$$

gdzie:  $Z$  – macierz standaryzowanych zmiennych pierwotnych,  $B$  – macierz współczynników głównych składowych,  $S$  – macierz głównych składowych,  $T$  – znak transpozycji.

Wyznaczone wartości na głównych składowych stanowiły podstawę do przeprowadzenia segmentacji rynku. Użyto w tym celu dwóch najczęściej stosowanych metod analizy skupień, tj. hierarchicznej metody Warda oraz niehierarchicznej metody k-średnich z przyjętą liczbą skupień wynikającą z metody Warda. Zasadność wyboru wskazanych metod potwierdzają liczne studia literaturowe poświęcone ocenie skuteczności procesów grupowania wielowymiarowych obiektów [Najman 2008].

Metoda Warda bazuje na analizie wariancji i zakłada łączenie tych obiektów, które powodują możliwie najmniejszy przyrost wariancji w skupieniu. Zgodnie z założeniami metody na każdym etapie grupowania  $k$  ( $k=1, \dots, n-1$ ) dla  $n-k$  istniejących skupień należy wyznaczyć sumę kwadratów odchyleń od średniej w obrębie każdego skupienia [Balicki 2009]. W sytuacji analizy jednowymiarowej (cecha  $X_j$ ) wariancja w skupieniu wynosi:

$$W_g^{(k)} = \sum_{i=1}^{n_g} (x_{ij} - \bar{x}_j)^2 \quad (g = 1, \dots, n - k) \quad (2)$$

Dla każdego możliwego podziału na danym etapie łączenia obliczany jest wskaźnik:

$$W^{(k)} = \sum_{g=1}^{n-k} W_g^{(k)} \quad (3)$$

gdzie:  $n_g$  – liczba elementów w skupieniu  $g$ ,  $x_{ij}$  – wartość cechy  $i$ -tej jednostki ( $i=1, \dots, n_g$ ) w skupieniu  $g$ .

W ostatnim etapie badania wykorzystano metodę k-średnich, której istota polega na optymalizowaniu pewnej funkcji jakości podziału obiektów. Jak zauważa Gatnar, funkcję jakości można zapisać w postaci formuły maksymalizującej ślad macierzy wariancji międzygrupowej lub minimalizującej ślad macierzy wariancji wewnątrzgrupowej [Gatnar 1998]. Zgodnie z istotą metody k-średnich algorytm postępowania zmierzający do uzyskania jak najbardziej podobnych obiektów wewnątrz danego skupienia sprowadza się do tego, że wstępnie tworzy się  $k$ -losowych skupień, a w kolejnych krokach przenosi się obiekty między skupieniami, tak aby minimalizować zmienność wewnątrz skupień i maksymalizować zmienność między skupieniami.

## WYNIKI BADAŃ

W tabeli 1 zaprezentowano wyniki analizy głównych składowych. W oparciu o kryterium stopnia wyjaśniania wariancji, które ustalono na poziomie 60%, wyodrębniono sześć głównych składowych. Interpretując zakres informacji przenoszonych przez poszczególne składowe, dokonano identyfikacji profili zakupowych występujących w próbie badanych respondentów. Cechy swoiste dla poszczególnych profili konsumentów przedstawiają się następująco:

1. Konsument (podatny na wpływy) kierujący się informacjami pochodzącymi z przekazów reklamowych, dla którego szczególne znaczenie posiadają różne formy reklamy i rekomendacje (składowa 1).
2. Konsument (eko klient) zorientowany na produkty ekologiczne o tradycyjnych recepturach, dla którego istotna jest moda, marka i opakowanie oraz możliwość degustacji wyrobów przed zakupem (składowa 2).
3. Konsument (dociekliwy ekspert) ceniący jakość, na którą składają się walory odżywcze, skład produktu, gatunek, brak konserwantów oraz niska zawartość tłuszczów (składowa 3).
4. Konsument (koneser) kierujący się smakiem, zapachem, świeżością i przydatnością do spożycia produktów (składowa 4).
5. Konsument (pragmatyk) zwracający uwagę na promocje handlowe, cenę, dostępność i wygląd (składowa 5).
6. Konsument (konserwatysta) o stałych preferencjach, kupujący produkty regionalne z określonego kraju (składowa 6).

Tabela 1. Macierz rotowanych składowych

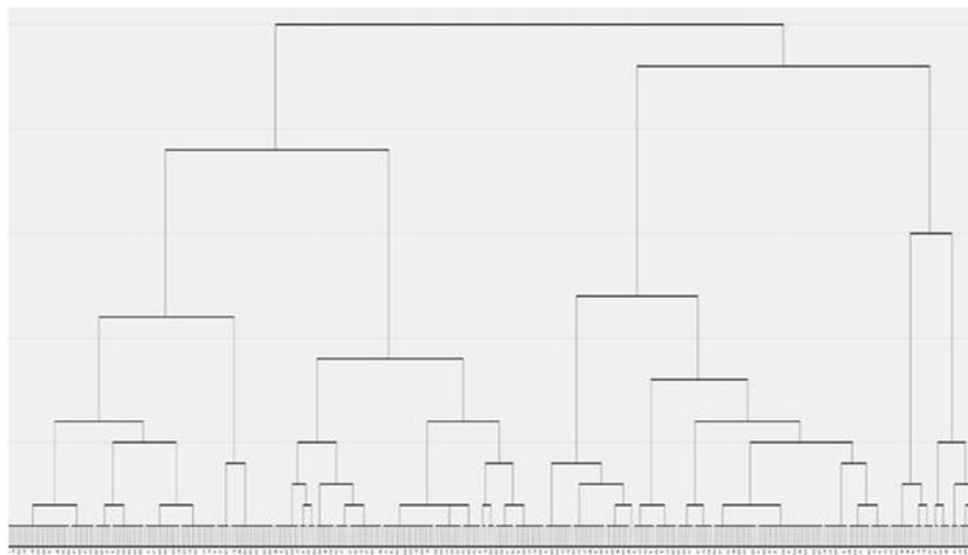
Czynniki wpływające na decyzje zakupowe	Składowa					
	1	2	3	4	5	6
Wygląd produktu	0,067	-0,150	0,260	0,399	0,417	0,063
Smak	-0,080	-0,060	0,075	0,637	0,205	0,001
Zapach	-0,020	-0,067	0,233	0,669	0,163	0,087
Gatunek	0,097	0,047	0,567	0,373	0,144	0,130
Skład produktu	-0,029	-0,064	0,796	0,187	0,113	0,143
Walory odżywcze	0,047	0,144	0,728	0,222	0,011	0,050
Brak konserwantów	0,083	0,223	0,766	0,005	-0,012	0,053
Świeżość produktu	0,004	0,082	0,179	0,712	0,055	-0,002
Niska zawartość tłuszczu	0,087	0,273	0,687	0,088	0,072	-0,066
Data przydatności do spożycia	0,087	0,169	0,100	0,619	-0,106	-0,038
Certyfikaty jakości	0,163	0,612	0,365	0,336	-0,226	0,007
Tradycyjne receptury	0,208	0,598	0,413	0,162	-0,186	0,016
Zdrowotność produktu	0,070	0,439	0,426	0,366	-0,110	0,115
Wygląd opakowania	0,148	0,660	-0,007	0,176	0,181	0,044
Wielkość opakowania	0,072	0,581	0,043	0,126	0,387	0,123
Marka produktów	0,116	0,646	0,122	0,023	0,202	0,198
Wyrób ekologiczny/bio	0,127	0,659	0,373	0,029	0,006	0,020
Cena produktu	0,077	0,101	0,126	0,071	0,697	-0,029
Rabaty cenowe	0,160	0,226	0,022	-0,030	0,742	-0,050
Dostępność w punkcie sprzedaży	0,092	0,325	-0,027	0,243	0,519	0,012
Degustacje w miejscu sprzedaży	0,221	0,702	0,094	-0,212	0,124	-0,016
Programy lojalnościowe	0,211	0,676	0,051	-0,138	0,135	0,016
Ekspozycja w miejscu sprzedaży	0,334	0,555	0,051	0,002	0,251	0,041
Promocje handlowe w miejscu sprzedaży	0,339	0,332	-0,191	0,144	0,441	0,143
Reklama telewizyjna	0,846	0,215	0,059	-0,089	0,141	0,034
Reklama radiowa	0,829	0,313	0,063	-0,105	0,086	0,065
Reklama w prasie	0,828	0,249	0,102	-0,035	0,160	0,006
Reklama na billboardach	0,856	0,166	-0,011	-0,025	0,160	0,054
Rekomendacje rodziny/znajomych	0,610	0,009	-0,016	0,359	-0,023	0,106

Czynniki wpływające na decyzje zakupowe	Składowa					
	1	2	3	4	5	6
Rekomendacje sprzedawcy	0,631	0,158	0,099	0,043	0,014	0,144
Opinie o produkcie/producentcie w mass mediach	0,663	0,090	0,094	0,128	0,010	0,334
Przyzwyczajenie	0,190	-0,058	-0,282	0,275	0,117	0,501
Kraj pochodzenia produktu	0,191	0,113	0,195	-0,047	-0,033	0,813
Producent z regionu	0,190	0,222	0,211	0,000	-0,049	0,816
Moda	0,383	0,436	0,047	-0,167	0,060	0,304

Źródło: opracowanie własne

Zgodnie z przyjętą metodyką badania liczba skupień została ustalona w oparciu o metodę Warda, przy wykorzystaniu odległości euklidesowych. Uzyskane wyniki prezentuje dendrogram (rysunek 1). Poziom odcięcia dendrogramu decydujący o liczbie skupień określono w oparciu o wskaźnik Calińskiego–Harabasz.

Rysunek 1. Diagram drzewkowy



Źródło: opracowanie własne

W kolejnym kroku wyodrębniono 8 skupień (segmentów) przy zastosowaniu metody k-średnich. Dla wyodrębnionych skupień obliczono przeciętne wartości na poszczególnych składowych (tabela 2). Uzyskane wyniki wskazują, że w segmencie 1 dominują konsumenci, których profile opisane są przez składowe 1 i 2, zaś decyzje zakupowe tej grupy nabywców są skrajnie różne od profilu reprezentowanego przez składową 4. W przypadku segmentu 2 podobnie jak dla segmentu 1 zachowania zakupowe nabywców z tej grupy najlepiej charakteryzowane są przez składową 1 i 2, co istotnie odróżnia tę grupę nabywców

od segmentu 3, dla którego kluczową rolę odgrywa wygląd, cena, dostępność i promocje handlowe. Jeżeli chodzi o segment 4, to można powiedzieć, że konsumenci z tego skupienia w swoich decyzjach zakupowych kierują się takimi cechami produktów, jak: walory odżywcze, skład, gatunek, brak konserwantów oraz niska zawartość tłuszczów. Dla nabywców zaliczonych do segmentów 5 i 7 duże znaczenie mają produkty regionalne z określonego kraju pochodzenia oraz przyzwyczajenie.

Z kolei, konsument zorientowany na produkty ekologiczne o tradycyjnych recepturach, dla którego istotne są moda, marka i opakowanie oraz możliwość degustacji wyrobów przed zakupem jest reprezentatywnym typem nabywcy dla segmentu 6. Ostatecznie konsumentów tworzących skupienie 8 najlepiej opisuje profil zakupowy nabywcy, dla którego wyznacznikami atrakcyjności oferty są smak, zapach, świeżość i termin przydatności do spożycia.

Tabela 2. Ostateczne centra skupień

	Skupienie							
	1	2	3	4	5	6	7	8
Składowa 1	0,437	0,908	-0,029	0,206	-0,214	-0,850	-0,830	-0,096
Składowa 2	0,540	0,504	-10,077	-0,998	0,118	0,979	-0,099	-0,396
Składowa 3	-0,773	0,230	0,093	0,425	0,358	0,514	-0,147	-10,972
Składowa 4	-20,642	0,194	-10,330	0,381	0,322	0,074	0,357	0,600
Składowa 5	-0,988	0,450	0,938	-0,268	-10,497	0,419	0,144	0,0531
Składowa 6	-0,134	0,327	0,272	-0,962	0,370	-0,997	10,031	-0,571

Źródło: opracowanie własne

Biorąc pod uwagę liczebności wyodrębnionych skupień, największy segment, tj. skupienie 2, tworzą nabywcy, których preferencje zakupowe najlepiej opisuje składowa 1 i 2 (tabela 3). Co istotne, dla tego segmentu wartości na pozostałych składowych są wyższe od wartości przeciętnych. Taka sytuacja może sugerować istnienie znacznej grupy nabywców nie poddających się jednoznacznej klasyfikacji. Z drugiej strony, segmentem, który można uznać za niszowy jest skupienie 1. Pomimo, że tę grupę nabywców można sprofilować podobnie jak skupienie 2 w oparciu o analizę wartości na 1 i 2 składowej, to należy podkreślić jej całkowity brak podobieństwa do charakterystyk zakupowych prezentowanych przez pozostałe składowe.

Tabela 3. Liczba obserwacji w każdym skupieniu

Skupienie	1	2	3	4	5	6	7	8	Suma
Liczebność	17	86	34	47	46	44	53	31	358

Źródło: opracowanie własne

## PODSUMOWANIE

Reasumując przedstawione w artykule rozważania na temat segmentacji rynku oraz przeprowadzone analizy statystyczne można sformułować następujące wnioski:

1. Właściwie przeprowadzona segmentacja rynku może stać się podstawą skutecznej strategii marketingowej na konkurencyjnym i ciągle zmieniającym się rynku.
2. Złożoność zachowań i postaw nabywców powodują, że należy poszukiwać coraz bardziej kompleksowych metod segmentacji rynku.
3. Analiza głównych składowych i metoda k-średnich mogą być uzupełnieniem dla metod segmentacji rynku wykorzystujących klasyczne kryteria podziału (demograficzne, społeczne i ekonomiczne).
4. Zastosowane metody statystyczne pozwoliły na wyodrębnienie względnie jednorodnych grup badanych nabywców na rynku mięsa i jego przetworów.
5. Zidentyfikowane grupy są na tyle różne między sobą, że można próbować opracowywać sprofilowane strategie marketingowe uwzględniające postawy i zachowania badanych nabywców na rynku.
6. Zaprezentowane badania i wnioski mogą stanowić podstawę do przeprowadzenia kolejnych, bardziej pogłębionych badań i analiz w celu poszukiwania efektywniejszych kryteriów segmentacji rynku.

## BIBLIOGRAFIA

- Balicki A. (2009) Statystyczna Analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne. Wydawnictwo Uniwersytetu Gdańskiego.
- Dębkowska K. (2010) Metody statystyczne w segmentacji rynku. *Economy and Management*, 4, 183-193.
- Garbarski L., Rutkowski I., Wrzosek W. (2006) Marketing. Punkt zwrotny nowoczesnej firmy. PWE, Warszawa.
- Gatnar E. (1998) Symboliczne metody klasyfikacji danych. PWN, Warszawa.
- Gatnar M., Walesiak E. (2004) Metody statystycznej analizy wielowymiarowej w badaniach marketingowych. AE, Wrocław.
- Han S., Ye Y., Chen Z. (2014) Category Role Aided Market Segmentation Approach to Convenience Store Chain Category Management. *Decision Support Systems*, 57, 296-308.
- Hunt S., Arnett D. (2014) Market Segmentation Strategy, Competitive Advantage, and Public Policy: Grounding Segmentation Strategy in Resource-Advantage Theory. *Australian Marketing Journal*, 12(1), 7-24.
- Kotler Ph. (2005) Marketing. Rebis, Poznań.
- Lipowski M., Angowski M. (2016) Gender and Consumer Behaviour in Distribution Channels of Services. *International Journal of Synergy and Research*, 5, 45-58.
- Maciejewski G. (2012) Konsument w strategii współczesnego przedsiębiorstwa. *Konsumpcja i Rozwój*, 2, 37-46.

- Mahajan D., Agarwal N., Agarwal A. (2008) Evolution of Market Segmentation. *Journal of Marketing Management*, 7 (2), 30-39.
- Maricic B., Dordevic A. (2015) Strategic Market Segmentation. *Marketing*, 46 (4), 243-251.
- Mazurek-Łopacińska K. (2005) *Badania marketingowe. Teoria i praktyka*. PWN, Warszawa.
- Muller H., Hamm U. (2014) Stability of Market Segmentation with Cluster Analysis – A Methodological Approach. *Food Quality and Preference*, 34, 70-78.
- Najman K. (2008) Symulacyjna analiza wpływu wyboru kryterium optymalności podziału obiektów na jakość uzyskanej klasyfikacji a algorytmach k-średnich. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, Taksonomia*, 15(7), 295-304.
- Reynolds T. (2006) Methodological and Strategy Development Implications of Decision Segmentation. *Journal of Advertising Research*, 46(4), 445-462.
- Stanimir A. (2006) *Analiza danych marketingowych. Problemy, metody, przykłady*. AE, Wrocław.
- Tonks D. (2009) Validity and the Design of Market Segments. *Journal of Marketing Management*, 25 (3-4), 341-356.
- Venter P., Wright A., Dibb S. (2015) Performing market segmentation: a performative perspective. *Journal of Marketing Management*, 31(1-2), 63-82.
- Verain M., Bartels J., Dagevos H., Sijtsema S., Onwezen M., Antonides G. (2012) Segments of Sustainable Food Consumers: a Literature Review. *International Journal of Consumer Studies*, 36, 123-132.
- Walesiak M. (2000) Segmentacja rynku. Kryteria i metody. [w:] Zeliaś A. (red.) *Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych*. Wyd. AE, Kraków, 191-201.
- Żakowska-Biemans S., Gutkowska K., Sajdakowska M. (2013) Segmentacja konsumentów z uwzględnieniem skłonności do zaakceptowania innowacji w produktach żywnościowych pochodzenia zwierzęcego. *Handel Wewnętrzny*, 4(345), 141-154.

#### APPLICATION OF K-MEANS METHODS IN SEGMENTATION OF BUYERS ON THE FOOD PRODUCTS MARKET

**Abstract:** Knowing buyers' preferences and behaviors is a key factor of contributing to success of every business. Effective marketing requires precise segmentation, according to which a personalized market offer is created that takes into account the capabilities and needs of a specific consumers group. The k-means method is a research method that can be the basis for segmentation of buyers as it reflects consumer responses to individual variants or attributes of products. The purpose of the publication is to present the possibilities of application of the k-means method in the segmentation procedure in the food market.

**Keywords:** Market segmentation, k-means method, buyer profiles



## OCENA PRAWDOPODOBIENSTWA REZYGNACJI Z POŚREDNICTWA URZĘDU PRACY W POSZUKIWANIU ZATRUDNIENIA

**Beata Bieszk-Stolorz (ORCID: 0000-0001-8086-9037)**

**Iwona Markowicz (ORCID: 0000-0003-1119-0789)**

Wydział Nauk Ekonomicznych i Zarządzania

Uniwersytet Szczeciński

e-mail: beatus@wneiz.pl; iwona.markowicz@wneiz.pl

**Streszczenie:** Częstymi przyczynami odpływu z rejestrów bezrobotnych jest odmowa podjęcia proponowanego zatrudnienia lub wykreślenie z powodu niestawiennictwa w wyznaczonym terminie. Celem artykułu jest analiza prawdopodobieństwa rezygnacji bezrobotnych z pośrednictwa urzędu pracy. W badaniu wykorzystano estymator Kaplana-Meiera. Umożliwił on ocenę prawdopodobieństwa ubywania jednostek z kohorty w kolejnych miesiącach w zależności od ich cech. Główną część badania poprzedzono analizą wskaźnikową umożliwiającą porównanie struktur osób wyrejestrowanych z powodu podjęcia pracy i osób wykreślonych z rejestru.

**Słowa kluczowe:** estymator Kaplana-Meiera, bezrobocie, rezygnacja z pośrednictwa urzędu pracy

### WSTĘP

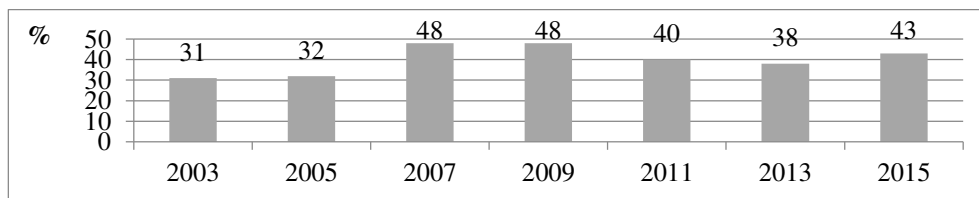
Analizy dotyczące rynku pracy skupiają się zazwyczaj na osobach wychodzących z bezrobocia poprzez podjęcie pracy. Dane statystyczne gromadzone w powiatowych urzędach pracy dostarczają bogatych informacji na temat innych powodów wyrejestrowań. Jest ich kilkadziesiąt. Przykładowo są to: przejście na rentę lub emeryturę, wyjazd za granicę na okres dłuższy niż 30 dni, zmiana miejsca zamieszkania, śmierć, przyznanie zasiłku przedemerytalnego. Częstymi przyczynami odpływu z rejestrów jest odmowa podjęcia proponowanego zatrudnienia lub wykreślenie z powodu niestawiennictwa w urzędzie pracy w wyznaczonym terminie. Przedmiotem niniejszego badania jest grupa osób, które zostały wykreślone z urzędu z nieokreślonej przyczyny. Kohortę stanowią osoby

zarejestrowane w 2013 roku w Powiatowym Urzędzie Pracy w Szczecinie (PUP), obserwowane do końca 2014 roku. Celem artykułu jest analiza prawdopodobieństwa rezygnacji bezrobotnych z pośrednictwa urzędu pracy w poszukiwaniu zatrudnienia w zależności od wybranych cech: płci, wykształcenia wieku, stażu pracy i liczby kolejnych zarejestrowań. Ze względu na występowanie obserwacji cenzurowanych w badaniu wykorzystano estymator Kaplana-Meiera. Umożliwił on ocenę prawdopodobieństwa ubywania jednostek z kohorty w kolejnych miesiącach. Zasadniczą część badania poprzedzono analizą wskaźnikową umożliwiającą porównanie struktury osób wyrejestrowanych z powodu podjęcia pracy i osób wykreślonych z rejestru. W badaniu wykorzystano dane indywidualne 22078 osób pochodzące z systemu SYRIUSZ.

## PROBLEM POZORNEGO BEZROBOCIA

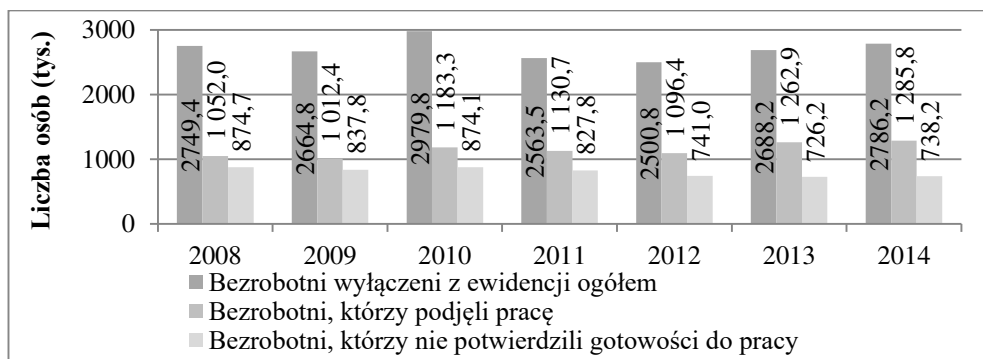
Wszystkie osoby zarejestrowane w urzędach pracy można podzielić na dwie grupy: prawdziwych i pozornych, czyli osoby, które nie są zainteresowane pracą (nie szukają jej i/lub nie są gotowi jej podjąć) oraz pracujących na czarno lub w inny sposób osiągających dochód nie mniejszy niż 1200 zł miesięcznie [Kotowska 2014, s. 197]. Stanowią oni duży odsetek wszystkich osób zarejestrowanych (rysunek 1).

Rysunek 1. Odsetek osób zarejestrowanych pozornie bezrobotnych w Polsce



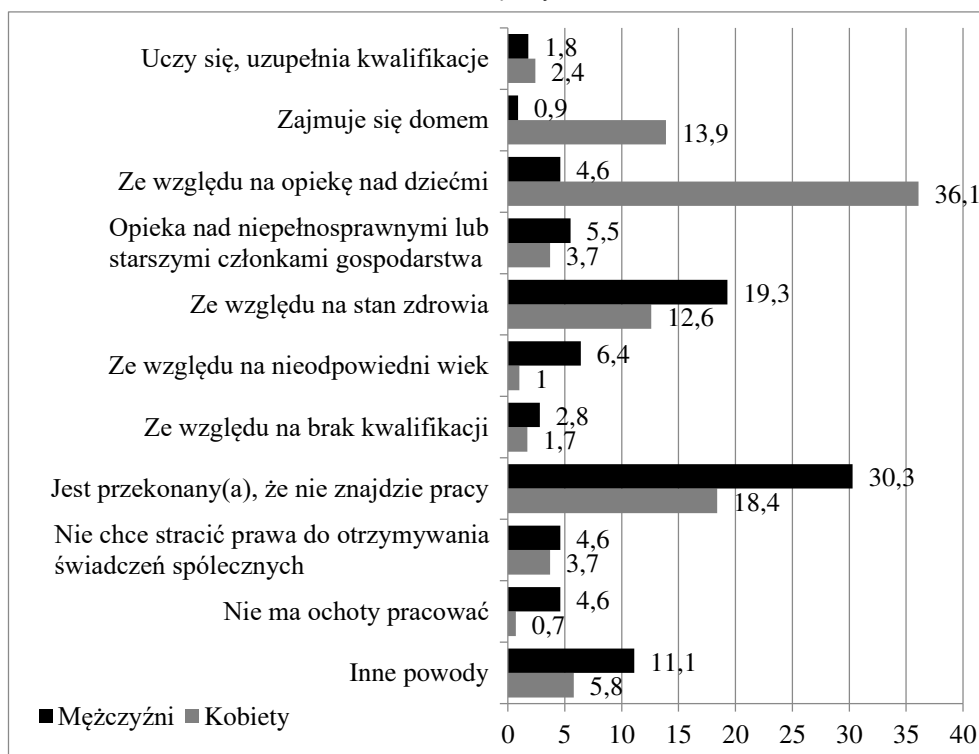
Źródło: opracowanie własne na podstawie: [Kotowska 2014], [Czapiński, Panek 2015]

Rysunek 2. Liczba wyrejestrowań z urzędów pracy w Polsce w latach 2008-2014 (w tys.)



Źródło: opracowanie własne na podstawie: Rocznik Statystyczny Pracy 2010, 2012, 2015

Rysunek 3. Przyczyny nieposzukiwania pracy wśród zarejestrowanych bezrobotnych w 2015 roku – odsetek kobiet i mężczyzn



Źródło: opracowanie własne na podstawie: [Czapiński, Panek 2015]

Wśród wielu powodów wyrejestrowania z urzędów pracy w Polsce na szczególną uwagę zasługuje wykreślenie z powodu braku gotowości do pracy. W latach 2008-2014 stanowiło ono od 27% do 32% wszystkich wyrejestrowań w Polsce i było ono niewiele niższe od podjęcia pracy (rysunek 2). Część osób bezrobotnych nie informuje urzędu o znalezieniu pracy sądząc, że obowiązek ten należy do ich pracodawcy. Formalnie powinni zrobić to w ciągu tygodnia. Urzędy dopóki nie wykreślą bezrobotnego płacą za niego składki. Próbą zmniejszenia skali tego zjawiska jest nałożenie kary na osoby wykreślone z rejestru z własnej winy. Jest nią utrudnione odzyskanie statusu bezrobotnego, a co za tym idzie prawa do ubezpieczenia zdrowotnego i zasiłku. Jeżeli osoba zarejestrowana odmówiła bez uzasadnionej przyczyny przyjęcia propozycji odpowiedniej pracy lub innej formy pomocy zostaje pozbawiona statusu bezrobotnego. Jeśli taka sytuacja miała miejsce po raz pierwszy, obowiązuje karencja 120 dni, zanim urząd ponownie wpisze taką osobę do rejestru i obejmie ubezpieczeniem. Natomiast jeżeli wykreślenie nastąpiło po raz drugi, to karencja wynosi 180 dni, a po oraz kolejny – 270 dni. Jeżeli osoba zarejestrowana nie stawiła się w powiatowym urzędzie pracy w wyznaczonym terminie, to pozbawienie statusu poszukującego pracy następuje

na okres 120 dni [Ustawa z dnia 20 kwietnia 2004 r. ...]. Bezrobotni zarejestrowani podają różne przyczyny nieposzukiwania pracy (rysunek 3). W 2015 roku kobiety najczęściej zaznaczały opiekę nad dziećmi, a mężczyźni wyrażali przekonanie, że nie znajdują pracy [Czapiński, Panek 2015, s. 415].

## DANE WYKORZYSTANE W BADANIU

W badaniu wykorzystano dane indywidualne 22 078 osób zarejestrowanych w Powiatowym Urzędzie Pracy w Szczecinie i obserwowanych do końca 2014 roku pochodzące z systemu SYRIUSZ. Dane te zawierały informacje o czasie zarejestrowania, płci, wieku, wykształceniu, stażu pracy, liczbie kolejnych zarejestrowań osoby bezrobotnej oraz przyczynie wyrejestrowania. Strukturę badanych bezrobotnych przedstawiono w tabeli 1.

Tabela 1. Struktura badanych osób bezrobotnych

Grupa		Razem		Praca		Wykreślenie	
		Liczba	%	Liczba	%	Liczba	%
Ogółem		22078	100,00	9678	43,84	8965	40,61
Płeć	Kobiety ( <i>K</i> )	9770	44,25	4836	49,50	3264	33,41
	Mężczyźni ( <i>M</i> )	12308	55,75	4842	39,34	5701	46,32
Wiek	18-24 ( <i>W</i> <sub>1</sub> )	4148	18,79	1506	36,31	2257	54,41
	25-34 ( <i>W</i> <sub>2</sub> )	7356	33,32	3614	49,13	2966	40,32
	35-44 ( <i>W</i> <sub>3</sub> )	4259	19,29	1869	43,88	1734	40,71
	45-54 ( <i>W</i> <sub>4</sub> )	3497	15,84	1642	46,95	1214	34,72
	55-59 ( <i>W</i> <sub>5</sub> )	2185	9,90	837	38,31	629	28,79
	60-64 ( <i>W</i> <sub>6</sub> )	633	2,87	210	33,18	165	26,07
Wykształcenie	Co najwyżej gimnazjalne ( <i>S</i> <sub>1</sub> )	5123	23,20	1410	27,52	2932	57,23
	Zasadnicze zawodowe ( <i>S</i> <sub>2</sub> )	5016	22,72	1968	39,23	2220	44,26
	Średnie ogólnokształcące ( <i>S</i> <sub>3</sub> )	2859	12,95	1226	42,88	1223	42,78
	Średnie zawodowe ( <i>S</i> <sub>4</sub> )	4086	18,51	1943	47,55	1415	34,63
	Wyższe ( <i>S</i> <sub>5</sub> )	4994	22,62	3131	62,70	1175	23,53
Staż	Bez doświadczenia zawodowego ( <i>D</i> <sub>0</sub> )	7557	34,23	2335	30,90	4195	55,51
	Z doświadczeniem zawodowym ( <i>D</i> <sub>1</sub> )	14521	65,77	7343	50,57	4770	32,85
Liczba zarejestrowań	Zarejestrowani po raz pierwszy ( <i>Z</i> <sub>0</sub> )	5418	24,54	2495	46,05	2045	37,74
	Zarejestrowani po raz kolejny ( <i>Z</i> <sub>1</sub> )	16660	75,46	7183	43,12	6920	41,54

Źródło: opracowanie własne

Urzędy pracy podają kilkadziesiąt powodów wyrejestrowania. Najbardziej liczną grupą powodów było szeroko pojęte podjęcie pracy (podjęcie zatrudnienia,

działalności gospodarczej lub pracy subsydiowanej przez urząd). W badanym okresie pracę podjęło prawie 44% bezrobotnych zarejestrowanych. Nie mniej liczną grupę (prawie 41%) stanowiły osoby, które nie zgłosiły się w urzędzie w wyznaczonym terminie lub nie przyjęły propozycji pracy (wykreślenie).

## METODA BADAWCZA

Zasadniczą część badania poprzedzono analizą wskaźnikową, której celem było porównanie wyrejestrowań z powodu podjęcia pracy z wykreśleniami z rejestru (tabela 1). W dalszej części skorzystano z metod z obszaru analizy trwania powszechnie stosowanych w demografii i naukach medycznych do badania czasu trwania życia ludzkiego. Można nimi również analizować czas trwania zjawisk społeczno-ekonomicznych: żywotność firm [Markowicz 2012], aktywność ekonomiczną ludności [Landmesser 2013], dynamikę ubóstwa w miejskich i wiejskich gospodarstwach domowych [Sączewska-Piotrowska 2016], ryzyko kredytowe [Matuszyk 2015, Wycinka, 2015], czas trwania bezrobocia [Bieszk-Stolorz 2013, Bieszk-Stolorz, Markowicz 2012]. Analizie podlega czas od momentu rozpoczęcia pewnego procesu do momentu wystąpienia zdarzenia, czyli czas trwania jednostki w danym stanie (czas działania firmy, czas bezrobocia, czas do zaprzestania spłaty kredytu). Przyjmuje się, że czas ten jest zmienną losową  $T$ . Podstawowym pojęciem w analizie trwania jest funkcja trwania, nazywana również funkcją przeżycia, zdefiniowana następująco:

$$S(t) = P(t > T) = 1 - F(t) \quad (1)$$

gdzie:

$T$  – czas trwania zjawiska,

$F(T)$  – dystrybuanta rozkładu zmiennej losowej  $T$ .

Ze względu na to, że nie jest znany rozkład czasu trwania w bezrobociu oraz wystąpiły obserwacje cenzurowane w badaniu wykorzystano estymator Kaplana-Meiera [Kaplan, Meier 1958]:

$$\hat{S}(t) = \prod_{j: t_j \leq t} \left( 1 - \frac{d_j}{n_j} \right) \quad (2)$$

gdzie:

$d_j$  – liczba zdarzeń w momencie  $t_j$ ,

$n_j$  – liczba jednostek narażonych na zdarzenie do momentu  $t_j$ .

Funkcja trwania informuje o tym, jakie jest prawdopodobieństwo niezajścia zdarzenia co najmniej do czasu  $t$ . Czasami wygodniej jest analizować dystrybuantę  $F(T)$ , która wyraża prawdopodobieństwo, że zdarzenie zajdzie najpóźniej do czasu  $t$ . Jeżeli zdarzeniem jest wykreślenie z rejestru, to estymator funkcji trwania informuje o tym, jakie jest prawdopodobieństwo pozostania w rejestrze z powodu niewykreślenia, a estymator dystrybuanty pozwala określić, jakie jest prawdopodobieństwo wykreślenia.

Analizowane zbiorowości można dzielić na grupy ze względu na badane cechy, oszacować funkcje trwania dla każdej z tych grup i zbadać istotności różnic między nimi. Ponieważ nie są znane rozkłady czasu trwania stosuje się testy nieparametryczne oparte na porządku rangowym czasów trwania. Nie ma niestety powszechnie akceptowanych metod wyboru testu w danej sytuacji. Większość z nich daje rzetelne wyniki tylko przy dużych próbach, natomiast efektywność testów przy małych próbach jest mniej poznana. Do porównania dwóch krzywych trwania często stosuje się test log-rank [Kleinbaum, Klein 2005 s. 57-61]. Służy on do weryfikacji hipotezy  $H_0: S_1(t) = S_2(t)$  o równości krzywych trwania wyznaczonych dla obu grup. Statystyka testowa jest porównywana przy danym poziomie istotności z rozkładem chi kwadrat o jednym stopniu swobody. Test ten ma największą moc, gdy różnica między funkcjami hazardu dla pojedynczych podgrup jest stała w czasie [Landmesser 2013 s. 48-51]. Wstępna analiza z wykorzystaniem funkcji  $\ln(-\ln S(t))$  oraz pewne ograniczenia wynikające z założeń dla innych testów potwierdziły zasadność wykorzystania w badaniu testu log-rank.

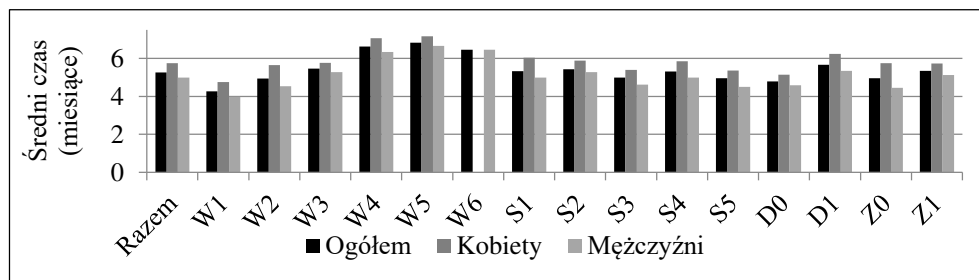
## WYNIKI BADAŃ

Wstępna analiza wskaźnikowa pozwoliła na zidentyfikowanie grup osób bezrobotnych najczęściej wykreślanych z rejestru. Mężczyźni częściej niż kobiety byli wykreśleni (46,32%), a kobiety częściej podejmowały pracę (49,50%). Największy odsetek wykreśleń był w grupie osób najmłodszych od 18 do 24 lat (54,41%) i zmniejszał się w kolejnych grupach wieku. Największy odsetek wyrejestrowań z powodu podjęcia pracy dotyczył osób w wieku 25-34 lata (49,13%). Odsetek osób podejmujących pracę rósł wraz z poziomem wykształcenia i najwyższy był dla osób z wykształceniem wyższym (62,70%). Najwięcej wykreśleń było wśród bezrobotnych z wykształceniem co najwyżej gimnazjalnym (57,23%). Odsetek wykreśleń malał wraz ze wzrostem poziomu wykształcenia. Posiadanie jakiegokolwiek doświadczenia zawodowego wpływało na częstość wyrejestrowań do pracy (50,57%), a jego brak – na częstość wykreśleń (55,51%). Osoby zarejestrowane po raz pierwszy częściej podejmowały pracę (46,05), a kolejny raz – częściej były wykreślane z rejestru (41,54%).

W dalszej części badania uwzględniono czas zarejestrowania osoby bezrobotnej w PUP. Obserwowano wszystkie osoby zarejestrowane w 2013 roku. Za zdarzenie kończące obserwację przyjęto fakt wykreślenia danej osoby z rejestru z powodu niezgłoszenia się w urządzie w wyznaczonym terminie lub z powodu nieprzyjęcia propozycji pracy. Za koniec okresu obserwacji przyjęto koniec 2014 roku. W badaniu wystąpiły dane cenzurowane. Były to obserwacje niezakończone wystąpieniem zdarzenia końcowego przez upływem 2014 roku lub obserwacje zakończone wyrejestrowaniem z powodu innego niż wykreślenie. Analiza średniego czasu do wyrejestrowania wykazała, że we wszystkich wyodrębnionych podgrupach mężczyźni szybciej niż kobiety byli wykreśleni z rejestru (rysunek 4),

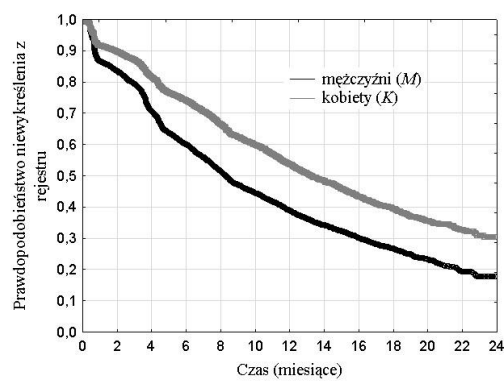
przy czym średni czas dla wszystkich bezrobotnych ogółem był równy 5,3 miesiąca.

Rysunek 4. Średni czas do wykreślenia według cech osób bezrobotnych (miesiące)



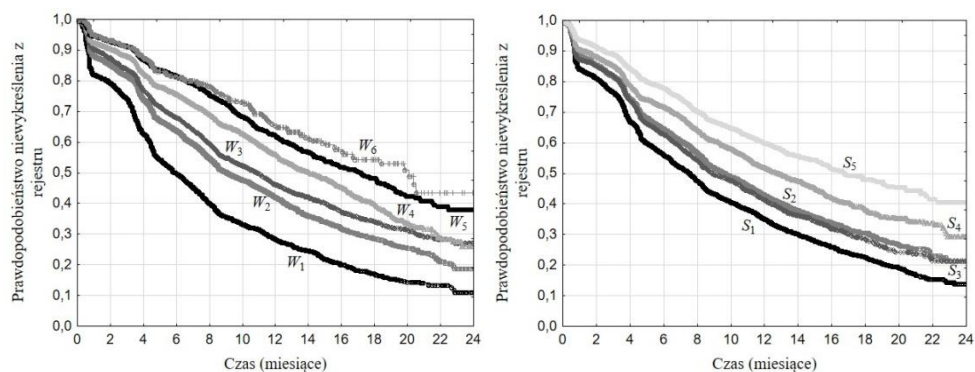
Źródło: opracowanie własne

Rysunek 5. Estymatory Kaplana-Meiera – wykreślenie według płci



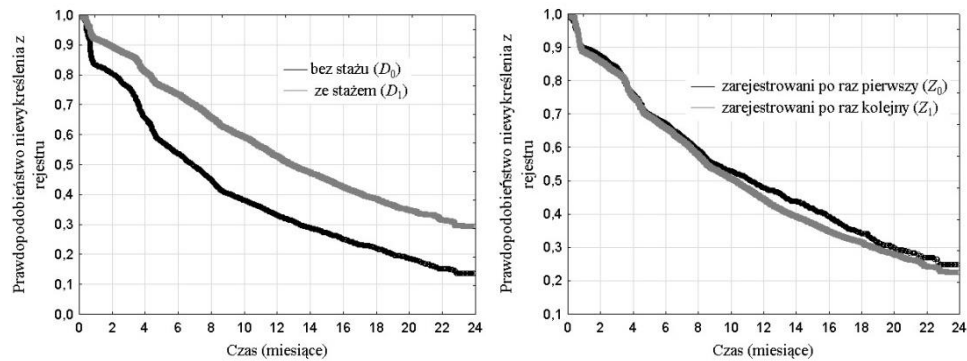
Źródło: opracowanie własne

Rysunek 6. Estymatory Kaplana-Meiera – wykreślenie według wykształcenia i grupy wieku



Źródło: opracowanie własne

Rysunek 7. Estymatory Kaplana-Meiera – wykreślenie według stażu pracy i liczby kolejnych zarejestrowań



Źródło: opracowanie własne

Następnie oceniono prawdopodobieństwo wyrejestrowania do czasu  $t$  osób bezrobotnych z powodu wykreślenia. Na rysunkach 5-7 przedstawiono estymatory Kaplana-Meiera wyznaczone dla wyodrębnionych w badaniu grup osób bezrobotnych. W opisie wyników przedstawiono wartości dopełnień do jedności tych estymatorów, czyli wyrażenie  $1 - \hat{S}(t)$ , gdyż taka interpretacja, ze względu na analizowane zjawisko, jest bardziej czytelna. Istotność różnic między parami krzywych trwania badano korzystając z testu log-rank (tabela 2).

Tabela 2. Wyniki testu log-rank dla par krzywych trwania w wyodrębnionych grupach cech osób bezrobotnych

Pary krzywych trwania	Test log-rank	Pary krzywych trwania	Test log-rank
$(K, M)$	-19,5584 ( $p = 0,0000$ )	$(W_4, W_6)$	-4,0675 ( $p = 0,0005$ )
$(W_1, W_2)$	-13,5694 ( $p = 0,0000$ )	$(W_5, W_6)$	-0,9830 ( $p = 0,3256$ )
$(W_1, W_3)$	-16,0771 ( $p = 0,0000$ )	$(S_1, S_2)$	-8,5546 ( $p = 0,0000$ )
$(W_1, W_4)$	-21,3550 ( $p = 0,0000$ )	$(S_1, S_3)$	-5,5923 ( $p = 0,0000$ )
$(W_1, W_5)$	-22,4636 ( $p = 0,0000$ )	$(S_1, S_4)$	-14,2753 ( $p = 0,0000$ )
$(W_1, W_6)$	-14,0922 ( $p = 0,0000$ )	$(S_1, S_5)$	-21,2524 ( $p = 0,0000$ )
$(W_2, W_3)$	-4,5637 ( $p = 0,0001$ )	$(S_2, S_3)$	1,4284 ( $p = 0,1532$ )
$(W_2, W_4)$	-11,4015 ( $p = 0,0000$ )	$(S_2, S_4)$	-6,3684 ( $p = 0,0000$ )
$(W_2, W_5)$	-14,4824 ( $p = 0,0000$ )	$(S_2, S_5)$	-13,2121 ( $p = 0,0000$ )
$(W_2, W_6)$	-9,2572 ( $p = 0,0000$ )	$(S_3, S_4)$	-6,7677 ( $p = 0,0000$ )
$(W_3, W_4)$	-6,9043 ( $p = 0,0000$ )	$(S_3, S_5)$	12,8871 ( $p = 0,0000$ )
$(W_3, W_5)$	-10,8377 ( $p = 0,0000$ )	$(S_4, S_5)$	-6,5452 ( $p = 0,0000$ )
$(W_3, W_6)$	-7,4171 ( $p = 0,0000$ )	$(D_0, D_1)$	-28,0469 ( $p = 0,0000$ )
$(W_4, W_5)$	-4,8862 ( $p = 0,0000$ )	$(Z_0, Z_1)$	1,9417 ( $p = 0,0522$ )

Źródło: opracowanie własne



Prawdopodobieństwo wykreślenia z rejestru urzędu było większe dla mężczyzn niż kobiet. Najszybciej wykreślane z rejestru były osoby młode w wieku od 18 do 24 lat, a najwolniej osoby w wieku przedemerytalnym 55-59 i 60-64. Test log-rank wskazał na brak istotnych różnic w przebiegu krzywych trwania dla obu tych grup. Również poziom wykształcenia determinował prawdopodobieństwo wykreślenia z rejestru – największe było w przypadku osób z wykształceniem co najwyżej gimnazjalnym, a najmniejsze z wykształceniem wyższym. Test log-rank potwierdził brak istotnych różnic w przebiegu krzywych trwania dla osób z wykształceniem zasadniczym zawodowym i średnim ogólnokształcącym. Osoby bez żadnego doświadczenia zawodowego były szybciej wykreślane z rejestru niż posiadające staż pracy. Test log-rank wskazał na brak istotnych różnic w przebiegu krzywych trwania dla osób pierwszy raz oraz kolejny raz zarejestrowanych w urzędzie. Wynika z tego, że ta cecha nie była determinantą czasu do wykreślenia z urzędu pracy.

## PODSUMOWANIE

Badanie przedstawione w artykule pozwoliło na zidentyfikowanie grupy osób wykreślonych z powodu niestawiennictwa w urzędzie w wyznaczonym terminie lub odmawiających propozycji przyjęcia zatrudnienia. Liczba wykreśleń z urzędów pracy z powodów leżących po stronie osoby bezrobotnej odzwierciedla problem pozornego bezrobocia. Krótki średni czas do wykreślenia (5,3 miesiąca) świadczy o dość szybkiej rezygnacji z pośrednictwa urzędów w poszukiwaniu pracy i niekorzystaniu z proponowanych różnorodnych form aktywizacji. Nie wiadomo, ile z tych osób znalazło pracę i tylko nie dopełniło formalności związanych z wyrejestrowaniem się, a ile podjęło pracę na czarno. Problem ten dotyczy przede wszystkim mężczyzn, ludzi młodych, po raz pierwszy zarejestrowanych w urzędzie pracy, z wykształceniem co najwyżej gimnazjalnym, nie posiadających doświadczenia zawodowego. Analiza wskaźnikowa i analiza prawdopodobieństwa wykreślenia z rejestru wskazały te same grupy zagrożone wykreśleniem. Test log-rank nie potwierdził różnic w czasie do wykreślenia w zależności od liczby zarejestrowań w urzędzie osoby poszukującej pracy. Przeprowadzona analiza dostarcza jeszcze jednego ważnego wniosku: uwarunkowania prawne pozwalające urzędom na stosowanie sankcji wobec takich osób nie przynoszą spodziewanych rezultatów.

## BIBLIOGRAFIA

- Bieszk-Stolorz B. (2013) Analiza historii zdarzeń w badaniu bezrobocia. Volumina.pl  
Daniel Krzanowski, Szczecin.
- Bieszk-Stolorz B., Markowicz I. (2012) Modele regresji Coxa w analizie bezrobocia.  
CeDeWu, Warszawa.

- Czapiński J., Panek T. (red.) (2015) *Diagnoza Społeczna 2015. Warunki i jakość życia Polaków*. Rada Monitoringu Społecznego, Warszawa.
- Kaplan E.L., Meier P. (1958) Non-Parametric Estimation from Incomplete Observations. *Journal of American Statistical Association*, 53, 457-481.
- Kleinbaum D., Klein M. (2005) *Survival Analysis. A Self-Learning Text*. Springer, New York.
- Kotowska I. E. (red.) (2014) *Rynek pracy i wykluczenie społeczne w kontekście percepcji Polaków*. Diagnoza społeczna 2013. MPiPS, Warszawa.
- Landmesser J. (2013) *Wykorzystanie metod analizy czasu trwania do badania aktywności ekonomicznej ludności w Polsce*. Wydawnictwo SGGW, Warszawa.
- Markowicz I. (2012) *Statystyczna analiza żywotności firm*. Wydawnictwo Naukowe Uniwersytetu Szczecińskiego.
- Matuszyk A. (2015) *Zastosowanie analizy przetrwania w ocenie ryzyka kredytowego klientów indywidualnych*. CeDeWu, Warszawa.
- Rocznik Statystyczny Pracy 2010, 2012, 2015 (2011, 2013, 2016), GUS, Warszawa.
- Sączewska-Piotrowska A. (2016) Dynamika ubóstwa w miejskich i wiejskich gospodarstwach domowych. *Wiadomości Statystyczne*, 7, 39-59.
- Ustawa z dnia 20 kwietnia 2004 r. o promocji zatrudnienia i instytucjach rynku pracy (Dz. U. z 2017 r. poz. 1065).
- Wycinka E. (2015) Modelowanie czasu do zaprzestania spłat rat kredytu lub wcześniejszej spłaty kredytu jako zdarzeń konkurujących. *Problemy Zarządzania*, 13(3), t. 2, 146-157.

#### **THE ASSESSMENT OF THE PROBABILITY OF RESIGNATION FROM INTERMEDIATION OF LABOUR OFFICE IN JOB SEEKING**

**Abstract:** Frequent reasons of deregistration is refusal of acceptance the proposed employment or erasure due to absence in the office in due time. The aim of the paper is analysis of probability of resignation of the unemployed people from the intermediation of the labour office. The Kaplan-Meier estimator was used. It allows the assessment of probability of outflow of units from the cohort in subsequent months with relation to their features. The main part of the research was preceded by the indicator analysis making comparison of the structure of persons deregistered because of taking a new job with the structure of persons erased from the registry possible.

**Keywords:** Kaplan-Meier estimator, unemployment, resignation from the intermediation of the labour office

## ROZKŁAD NORMALNY STÓP ZWROTU Z AKCJI WCHODZĄCYCH W SKŁAD NASTĘPUJĄCYCH INDEKSÓW GIEŁDOWYCH: WIG20, MWIG40 I SWIG80

**Krzysztof Borowski**

Kolegium Zarządzania i Finansów  
Szkoła Główna Handlowa w Warszawie  
e-mail: kborow@sgh.waw.pl

**Streszczenie:** W artykule zweryfikowana została teza o normalności rozkładu stóp zwrotu cen akcji komponentów indeksów giełdowych: WIG20, mWIG40 i sWIG80 w okresie od pierwszego notowania danego waloru (jednak nie wcześniej niż wprowadzenie systemu WARSET na GPW, tj. w dniu 17.10.2017 r.) do dnia 31.03.2017 r., dla stóp zwrotu w ujęciu: zamknięcie-zamknięcie, otwarcie-otwarcie, otwarcie-zamknięcie i overnight. Z wykorzystaniem testów Jarque-Bera, Shapiro-Wilka i D'Agostino-Pearsona podjęto próbę stworzenia rankingu spółek ze względu na zbieżność rozkładu ich stóp zwrotu do rozkładu normalnego.

**Słowa kluczowe:** rozkład normalny, stopy zwrotu, indeksy giełdowe, ranking spółek

### WSTĘP

Przez wiele lat w szeregu teoriach ekonomicznych przyjmowano, że rozkład stóp zwrotu na rynkach finansowych jest zbliżony do rozkładu normalnego. Tego typu podejście występuje m.in. w takich teoriach jak: teoria portfelowa Markowitza, model wyceny dóbr kapitałowych CAPM oraz model wyceny opcji Blacka-Scholesa. Jednak jak pokazały liczne prace empiryczne, rozkłady stóp zwrotu różnych instrumentów finansowych, różnią się od rozkładu normalnego. Niewątpliwie fakt ten podważa zasadność stosowania wymienionych wcześniej modeli i nakazuje poszukiwanie innych typów rozkładów stóp zwrotu, charakteryzujących się grubszymi ogonami niż rozkłady gaussowskie, które pozwalałyby lepiej modelować empiryczne rozkłady stóp zwrotu z instrumentów finansowych [Piasecki, Tomasik 2013].

Celem artykułu jest zweryfikowanie hipotezy o normalności rozkładów stóp zwrotu cen akcji spółek, wchodzących w skład następujących indeksów giełdowych: WIG20, mWIG40 i sWIG80 w okresie od pierwszego notowania danej spółki na GPW w Warszawie, do dnia 31.03.2017 r. Jeśli jednak data pierwszego notowania spółki jest wcześniejsza niż data wprowadzenia systemu WARSET (17.10.2000), wtedy jako datę pierwszego notowania danej spółki na GPW przyjmuje się datę implementacji systemu WARSET. Przed wprowadzeniem systemu notowań WARSET można było mówić jedynie o częściowych notowaniach ciągłych. Dopiero przejście do notowań cen instrumentów finansowych w systemie WARSET zbliżyło system notowań na GPW do systemów wykorzystywanych na rynkach rozwiniętych. Początkowa data notowań danego waloru została zamieszczona w tabelach 6, 7 i 8.

## PRZEGLĄD LITERATURY

Jedną z pierwszych prac poświęconych rozkładowi stóp zwrotu na rynkach finansowych była praca Bacheliera [Bachelier 1900], w której autor za pomocą modelu błędzenia losowego cen instrumentów finansowych, udowodnił, że różnice ceny w kolejnych okresach są zmiennymi o rozkładzie normalnym. Wartość oczekiwana tego rozkładu jest równa zero, a wariancja stanowi funkcję zależną od długości okresu. Pogłębione badania szeregów czasowych zostały przeprowadzone przez Kendalla [Kendall 1953], który na bazie tygodniowych stóp zwrotu akcji z rynku brytyjskiego wykazał normalność ich rozkładu ale też i lepkokurtyczność. Kolejną przełomową pracą był artykuł Osborne'a [Osborne 1959], który wykazał normalność rozkładu logarytmów naturalnych stóp zwrotu dla spółek notowanych na American Stock Exchange i New York Stock Exchange. Spośród bardziej współczesnych prac warto wspomnieć o artykule Scalasa i Kima [Scalas, Kim 2007], którzy za pomocą rozkładu stabilnego dokonali aproksymacji dziennych stóp zwrotu dla indeksów DJIA i MIBTEL. W przypadku tego pierwszego indeksu, testy zgodności Kołmogorowa i chi-kwadrat potwierdziły, a przypadku drugiego indeksu – zaprzeczyły hipotezie, że stopy zwrotu z indeksu mogą zostać przybliżone za pomocą rozkładu stabilnego. W pracy Barunika i in. [Barunik i in. 2010], hipoteza o normalności rozkładów stóp zwrotu dla indeksów WIG, PX i BUX w okresie marzec 2005 - marzec 2009, została odrzucona. W badaniach Ghahfarokhi i Ghahfarokhi [Ghahfarokhi, Ghahfarokhi 2009], wykazane zostało, że w przypadku stóp zwrotów indeksów: CAC40, DAX, DJAC, FTSE 100, ISEQ i S&P 500 wyznaczone z wykorzystaniem rozkładu stabilnego miary *Value at Risk* (VaR) są bliższe wielkościom rzeczywistym, niż miary otrzymane z wykorzystaniem rozkładu normalnego czy też t-studenta.

W przypadku badań poświęconych rynkowi polskiemu Bołt i Miłobędzki [Bołt, Miłobędzki 1994], analizując stopy zwrotu dla indeksu WIG oraz 21 akcji notowanych na GPW w okresie 1991-1993, doszli do wniosku, że w przeważającej mierze nie są to rozkłady normalne. Z kolei Fiszeder [Fiszeder 2000]

przeprowadził badanie rozkładu stóp zwrotu indeksu WIG oraz 12 innych indeksów światowych w okresie 02.01.1997-30.06.1999 r. za pomocą następujących testów zgodności: Pearsona, Kołmogorowa-Lillieforsa i Shapiro-Wilka. Dwa pierwsze testy pozwoliły na odrzucenie hipotezy zerowej o normalności rozkładu stóp zwrotu dla wszystkich badanych indeksów. W przypadku testu Shapiro-Wilka stwierdzono brak podstaw do odrzucenia hipotezy o normalności rozkładu stóp zwrotu jedynie w przypadku indeksu NIKK. Odrzucenie hipotezy o normalności rozkładów stóp zwrotu dla indeksu WIG20 w okresie 13.09.1997-15.02.2000 zostało udowodnione w pracy Rokity [Rokita 2000], a potwierdzone przez Osińską [Osińska 2006], która dla okresu: styczeń 1999 – lipiec 2001 przeanalizowała stopy zwrotu z indeksu WIG, WIG20 oraz dla 18 spółek-komponentów tego ostatniego indeksu. Również w pracy Witkowskiej i Kompy [Witkowska, Kompa 2007], rozkłady stóp zwrotu dla 12 spółek i 2 indeksów GPW w Warszawie, badane w okresie 02.01.2002-31.12.2005 okazały się nie być rozkładami normalnymi.

## METODOLOGIA BADANIA

Artykuł podzielony został na 4 części.

W części pierwszej przeprowadzona została weryfikacja hipotezy o normalności stóp zwrotu spółek wchodzących w skład następujących indeksów giełdowych: WIG20 (20 spółek), mWIG40 (40 spółek) i sWIG80 (80 spółek). Wszystkie analizowane spółki stanowiły komponenty indeksów giełdowych na dzień 31.03.2017 r. Jako pierwszy dzień notowania każdej ze spółek przyjęta została data jej debiutu giełdowego, pod warunkiem, że przypadła ona po dacie wprowadzenia systemu WARSET na GPW w Warszawie, tj. 17.10.2000 r. Dla spółek, których data debiutu giełdowego jest wcześniejszą od dnia 17.10.2000 r. przyjęta została data ich pierwszego notowania w dniu 17.10.2010 r. W badaniu analizie poddane zostały następujące stopy zwrotu:

- zamknięcie - zamknięcie (C-C):  $\frac{C_t - C_{t-1}}{C_{t-1}}$ ,
- otwarcie – otwarcie (O-O):  $\frac{O_t - O_{t-1}}{O_{t-1}}$ ,
- otwarcie – zamknięcie (O-C):  $\frac{C_t - O_t}{O_t}$ ,
- śródsesyjna (Overnight lub OV):  $\frac{O_t - C_{t-1}}{C_{t-1}}$ .

gdzie:  $C_t$  – cena zamknięcia na sesji  $t$ ,  $O_t$  – cena otwarcia na sesji  $t$ ,  $C_{t-1}$  – cena zamknięcia na sesji  $t-1$ ,  $O_{t-1}$  – cena otwarcia na sesji  $t-1$ ,

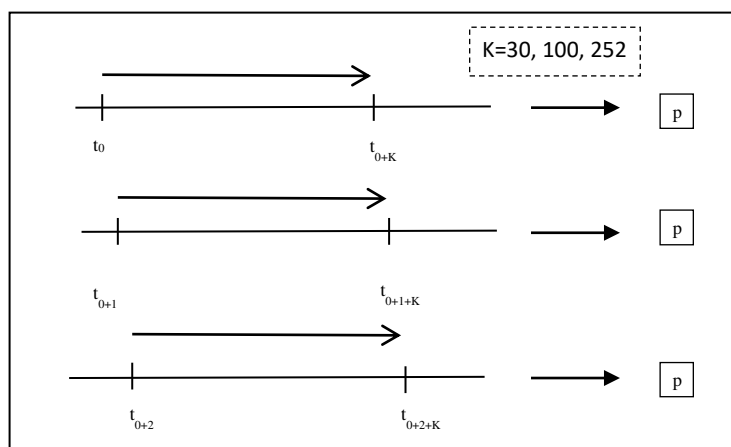
Wybór powyższych stóp zwrotu wynika z dwu przesłanek. Pierwszą jest przesłanka inwestycyjna – otwarcie lub zamknięcie transakcji ma miejsce w ściśle określonym momentach sesji po cenach otwarcia lub zamknięcia. Druga jest pochodną wcześniejszych artykułów naukowych, bowiem większość prac badawczych koncentruje się tylko i wyłącznie na stopach zwrotu zamknięcie-

zamknięcie. Autorowi nie są znane prace naukowe poświęcone problematyce normalności rozkładu stóp zwrotu na rynku polskim, innych niż stopy zwrotu zamknięcie-zamknięcie.

Weryfikacja hipotez statystycznych została przeprowadzona za pomocą pięciu następujących testów statystycznych: Jarque-Bera, Lillieforsa, Cramera von Misesa, Watsona oraz Andersona-Darlinga. Dla każdego z testów statystycznych obliczona zostanie wartość parametru  $p$  ( $p$  value). W przypadku, gdy wartość  $p$  jest większa lub równa od 0,05 – nie ma podstaw do odrzucenia hipotezy zerowej. W przeciwnym bowiem razie, tj. gdy wartość parametru  $p$  jest mniejsza niż 0,05; hipoteza zerowa zostaje odrzucona na rzecz hipotezy alternatywnej, która brzmi w sposób następujący: rozkład stóp zwrotu badanego instrumentu nie jest rozkładem normalnym.

W części drugiej przeprowadzona została weryfikacja hipotezy o normalności rozkładu stóp zwrotu dla 3 spółek z indeksu WIG20, tj. KGHM, Pekao S.A. i PKN Orlen w rocznych horyzontach czasowych, tj. dla lat: 2001- 2016. Celem tego badania jest wykazanie, że rozkład stóp zwrotu poszczególnych spółek może być normalny w krótszych interwałach czasowych. Część druga badania stanowi wprowadzenie do części trzeciej.

Rysunek 1. Wyznaczanie parametru  $p$  w  $K$  sesyjnym horyzoncie czasowym



Źródło: opracowanie własne

W trzeciej części badania weryfikacja hipotezy o normalności rozkładu stóp zwrotu przeprowadzona została według następującego schematu. Parametr  $p$  obliczono w chwili  $K$ , tj. dla pierwszych  $K$  sesji notowania akcji spółki na GPW w Warszawie. Jeśli  $t_0$  oznacza datę pierwszego notowania akcji na GPW w Warszawie, wtedy parametr  $p$  kalkulowany jest dla następujących sesji:  $t_0, t_{0+1}, t_{0+2}, \dots, t_{0+K}$ . Następny parametr  $p$  został wyznaczony dla  $K$ -sesyjnego horyzontu czasowego, ale przesuniętego o jedną sesję, tj. dla sesji odbywających się w chwilach:  $t_{0+1}, t_{0+2}, \dots, t_{0+K+1}$ . W podobny sposób kalkulowane będą wartości

parametru  $p$  dla pozostałych ciągów  $K$  sesji, tj. aż do momentu, kiedy ostatnia sesja w ciągu, przypada w dniu 31.03.2017 r. Dla wszystkich spółek wchodzących w skład analizowanych indeksów wartości  $p$  zostaną ustalone z wykorzystaniem testów: Jarque-Bera, Shapiro-Wilka i D'Agostino-Pearsona (pierwszy stopień swobody), a także dla trzech różnych długości horyzontu czasowego  $K$ : 30, 100 i 252 sesje (drugi stopień swobody) i stóp zwrotu: C-C, O-O, O-C i Overnight (trzeci stopień swobody). Kolejnym krokiem było podanie statystyki dla każdej z analizowanych spółek, która w szczególności obejmowała analizę liczby przypadków, dla danej wartości  $K$  i danego typu testu, kiedy nie było podstaw do odrzucenia hipotezy zerowej. W wyniku tego zabiegu otrzymuje się informację o częstości spełnienia warunku  $p > 0,05$ ; dla każdej z badanych stóp zwrotu, dla różnych  $K$  i dla różnych testów statystycznych.

Z uwagi na fakt, iż parametr  $p$  może być traktowany jako prawdopodobieństwo tego, że analizowany rozkład jest rozkładem normalnym, a więc im wyższa jest jego wartość, tym dany rozkład jest bardziej zbliżony do rozkładu normalnego, parametr  $p$  może być wykorzystany dla stworzenia rankingu spółek ze względu na bliskość rozkładu stóp zwrotu danej spółki w stosunku do rozkładu normalnego. Taki ranking spółek sporządzony został dla komponentów wchodzących w skład danego indeksu, określonej stopy zwrotu (C-C, O-O, C-O i *Overnight*), danej wartości  $K$  i danego typu testu statystycznego. W dalszej kolejności dla każdej ze spółek wchodzących w skład określonego indeksu, danej stopy zwrotu, danej wartości  $K$ , obliczona została suma rankingu danej spółki ze względu na stosowanie wszystkich rodzajów testów (ranking dla danej stopy zwrotu i danego  $K$ ):

$$S_{I+II+III} = S_I + S_{II} + S_{III} \quad (1)$$

gdzie:

- $S_I$  – miejsce w rankingu danej spółki z wykorzystaniem testu Jarque-Bera,
- $S_{II}$  – miejsce w rankingu danej spółki z wykorzystaniem testu Shapiro-Wilka,
- $S_{III}$  – miejsce w rankingu danej spółki z wykorzystaniem testu D'Agostino-Pearsona.

W ten sposób otrzymuje się 12 rankingów: (4 stopy zwrotu: C-C, O-O, O-C i *Overnight*) x (3 długości horyzontu inwestycyjnego  $K$ : 30, 100 i 252 sesje). W następnej kolejności, skalkulowana została suma miejsc danej spółki w 12 otrzymanych rankingach:

$$S_{I+\dots+XII} = \overset{K=30}{C-C}S + \overset{K=30}{O-O}S + \overset{K=30}{O-C}S + \overset{K=30}{OV}S + \overset{K=100}{C-C}S + \overset{K=100}{O-O}S + \overset{K=100}{O-C}S + \overset{K=100}{OV}S + \overset{K=252}{C-C}S + \overset{K=252}{O-O}S + \overset{K=252}{O-C}S + \overset{K=252}{OV}S \quad (2)$$

gdzie:

$\overset{K}{X}S$  – pozycja danej spółki w rankingu dla określonej długości  $K$  ( $K=30, 100$  i  $252$ ) i rodzaju stopy zwrotu  $X$  (C-C, O-O, O-C i *Overnight*).

Otrzymane w ten sposób dla każdej spółki sumy  $S_{I+\dots+XII}$  zostały wykorzystane do budowy rankingu globalnego dla każdego z indeksów.

W ostatniej części badania, na podstawie statystyk obliczonych w części trzeciej badania, skalkulowana została wartość średnia, maksymalna i minimalna każdej ze statystyk, kiedy  $K$  przybiera wartości 30, 100 i 252 sesje i dla każdego typu testu statystycznego.

Hipoteza główna postawiona w pracy została sformułowana w sposób następujący: w długim interwale czasowym rozkłady stóp zwrotu akcji wchodzących w skład indeksów giełdowych WIG20, mWIG40 i sWIG80 nie są rozkładami normalnymi. Przez długi interwał czasowy rozumie się horyzont inwestycyjny obejmujący kilka lat. Z kolei hipoteza pomocnicza pracy brzmi: w krótszych horyzontach inwestycyjnych, rozkłady stóp zwrotu akcji wchodzących w skład głównych indeksów GPW mogą być rozkładami normalnym. Hipotezę pomocniczą można także zapisać w nieco inny sposób: stopy zwrotu akcji wchodzących w skład głównych indeksów GPW są odcinkami normalne.

## WYNIKI I DYSKUSJA

### **Weryfikacja hipotezy o normalności rozkładu stóp zwrotu dla spółek wchodzących w skład indeksów: WIG20, mWIG40 i sWIG80**

W przypadku wszystkich spółek z indeksów WIG20, mWIG40 i sWIG80 analiza stóp zwrotu w okresie od daty pierwszego notowania danej spółki do dnia 31.03.2017 r. prowadziła do odrzucenia hipotezy zerowej na rzecz hipotezy alternatywnej. We wszystkich analizowanych przypadkach wartości parametru  $p$ , dla wszystkich stóp zwrotu, dla testów Jarque-Bera, Lillieforsa, Cramera Von Misesa, Watson i Andersona-Darlinga były równe zero z dokładnością do czterech miejsc po przecinku. Jedynie dla następujących spółek, wchodzących w skład indeksu sWIG80: CLNPHARMA i Stelmet oscylowały one w zakresie od 0,0001 do 0,001, co wynikało głównie z krótkiego okresu notowania tych firm na GPW w Warszawie (data pierwszego notowania tych spółek to odpowiednio: 01.12.2016 i 25.10.2016 r.).

### **Weryfikacja hipotezy o normalności rozkładu stóp zwrotu w poszczególnych latach dla wybranych spółek z WIGu20**

W czasie przeprowadzania testów statystycznych okazywało się, że zawężenie okresu analizy prowadziło do uzyskiwania wyższych wartości parametru  $p$ , co było równoznaczne w wielu przypadkach z brakiem konieczności odrzucania hipotezy zerowej. W związku z powyższym przeprowadzone zostało badanie normalności rozkładu stóp zwrotu w poszczególnych latach (2001-2016) dla trzech spółek z indeksu WIG20: Pekao S.A., PKN Orlen i KGHM (spółki te należą do pierwszej piątki spółek o największej kapitalizacji na GPW). Wyniki przedstawiono w tabelach 1-3. Skrótów w tabelach oznaczają rodzaj zastosowanego testu: J-B (Jarque-Bera), L (Lilliefors), CVM (Cramer von Mises), W (Watson), A-D (Anderson-Darling). W przypadku testu Lillieforsa wartość 0,1 oznacza wartość większą niż 0,1 (co w programie Eviews sygnalizowane jest jako  $>0,1$ ).



Czcionką pogrubioną zaznaczone zostały przypadki, dla których  $p > 0,05$ . Analiza otrzymanych wyników pozwala na wyciągnięcie wniosku, że w poszczególnych latach nie było podstaw do odrzucenia hipotezy zerowej (np. dla KGHM: rok 2005, 2009 i 2010 w układzie C-C) lub też hipoteza ta została odrzucona na rzecz hipotezy alternatywnej (np. PKN Orlen: rok 2004, układ C-C). O ile w przypadku stóp zwrotu C-C, O-O, O-C można wskazać wiele lat, kiedy nie było podstaw do odrzucenia hipotezy zerowej, o tyle w przypadku stóp zwrotu *overnight*, takich sytuacji nie ma, pomijając dla spółki PKN Orlen trzy przypadki parametru  $p$  o wartości wyższej niż 0,05 uzyskane przy zastosowaniu testu Jarque-Bera dla 2001, 2003 i 2009 r., jednak wykorzystanie innych testów statystycznych, tj. Lillieforsa, Cramera von Misesa, Andersona-Darlinga prowadziło do odrzucenia hipotezy zerowej na rzecz hipotezy alternatywnej.

Podsumowanie uzyskanych wyników zostało zamieszczone w tabeli 4, a czcionką pogrubioną zaznaczone zostały przypadki, kiedy to nie było podstaw do odrzucenia hipotezy zerowej przy użyciu co najmniej dwu różnych testów. Rozkład stóp zwrotu w poszczególnych latach, najbardziej zbliżony do rozkładu normalnego był udziałem PKN Orlen, przed Pekao S.A. i KGHM. W przypadku wszystkich trzech spółek, w żadnym z analizowanych lat, rozkład stóp zwrotu *overnight* nie był rozkładem normalnym, za wyjątkiem przypadku PKN Orlen omówionego wcześniej. Podobne badania mogą być przeprowadzone także dla poszczególnych fal wzrostowych i spadkowych widocznych na wykresach ww. spółek.

Tabela 1. Wyniki testów normalności rozkładów stóp zwrotu dla KGHM w latach 2001-2016

		2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
C-C	J-B	0,0001	0	0,03671	0	<b>0,0751</b>	0,0003	0	0	0,0196	<b>0,0776</b>	0	0	0	0,0289	0	0,0146
	L	0,0429	<b>0,1</b>	0,0211	0,0001	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	0,0007	<b>0,1</b>	<b>0,1</b>	0,0003	0,0003	0,0149	0,0096	0,0121	<b>0,1</b>
	CVM	0,0328	<b>0,0746</b>	0,0339	0	<b>0,1234</b>	<b>0,0541</b>	<b>0,0816</b>	0	<b>0,1205</b>	<b>0,1031</b>	0	0	0,0172	0,0006	0,0036	<b>0,105</b>
	W	0,0275	<b>0,064</b>	0,0343	0	<b>0,1117</b>	<b>0,0516</b>	<b>0,0632</b>	0	<b>0,1446</b>	<b>0,0821</b>	0	0	<b>0,221</b>	0,0003	0,0022	<b>0,1302</b>
	A-D	0,0334	0,045	0,044	0	<b>0,1191</b>	<b>0,0651</b>	0,0398	0	<b>0,0858</b>	<b>0,1303</b>	0	0	0,0033	0,0004	0,0009	<b>0,0672</b>
O-O	J-B	<b>0,0512</b>	0	0	0,0091	<b>0,5751</b>	0	0	0	0,0039	0	0	0	0,0011	0	0	0
	L	<b>0,1</b>	0,0217	0,0001	0,0014	<b>0,1</b>	0,0055	<b>0,0549</b>	0	<b>0,0782</b>	<b>0,1</b>	0,0004	0,0461	0,0003	0,0004	0,0035	0,0005
	CVM	<b>0,0561</b>	0,0038	0,0005	0,0006	<b>0,3257</b>	0,0002	0,0225	0	<b>0,2293</b>	<b>0,1715</b>	0	0,0002	0,0001	0,0001	0	0
	W	<b>0,461</b>	0,0028	0,0007	0,0003	<b>0,2901</b>	0,0001	0,0222	0	<b>0,221</b>	<b>0,1733</b>	0	0,0001	0,0001	0,0001	0	0
	A-D	0,036	0,0014	0,0004	0,0004	<b>0,3688</b>	0	0,0148	0	<b>0,1142</b>	<b>0,1537</b>	0	0	0	0,0001	0	0
O-C	J-B	<b>0,36</b>	0	<b>0,1161</b>	0,0004	<b>0,9215</b>	0	<b>0,5731</b>	0	0	0,0003	0	0,0004	0	0,0165	0	0
	L	<b>0,1</b>	0,0044	0,0255	0,0037	<b>0,1</b>	0,0176	<b>0,1</b>	0,0001	0,0357	<b>0,1</b>	0,0001	<b>0,1</b>	0,001	0,0105	0,0038	<b>0,0645</b>
	CVM	<b>0,2693</b>	0,0003	0,0378	0,0001	<b>0,5286</b>	0,0008	<b>0,4184</b>	0,0001	0,0418	<b>0,0972</b>	0	<b>0,4145</b>	0,0005	0,0013	0,0001	<b>0,0707</b>
	W	<b>0,2723</b>	0,0002	0,031	0	<b>0,5286</b>	0,0004	<b>0,4015</b>	0	0,0357	<b>0,0788</b>	0	<b>0,3743</b>	0,0009	0,0007	0,0006	<b>0,0663</b>
	A-D	<b>0,2079</b>	0,0003	0,048	0	<b>0,6413</b>	0,0006	<b>0,2285</b>	0	0,0196	0,0417	0	<b>0,287</b>	0,0004	0,0001	0,0007	0,0076
Overnight	J-B	0	0	0,0095	0	0	0	0	0	0	0	0	0	0	0	0	0
	L	0,0007	0	0	0	0	0	0	0	0,0003	0,0006	0	0	0,0001	0	0	0
	CVM	0,0003	0	0,0001	0	0	0	0	0	0,0094	0,0017	0	0	0	0	0	0
	W	0,0002	0	0,0001	0	0	0	0	0	0,0068	0,0009	0	0	0	0	0	0
	A-D	0,0001	0	0,0002	0	0	0	0	0	0,0065	0,0021	0	0	0	0	0	0

Źródło: opracowanie własne

Tabela 2. Wyniki testów normalności rozkładów stóp zwrotu dla Pekao S.A. w latach 2001-2016

		2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
C-C	J-B	0,0024	0,0214	<b>0,3849</b>	0	<b>0,0529</b>	0,0005	0	<b>0,3625</b>	0	<b>0,2832</b>	0	0,045	0	0,0192	<b>0,251</b>	0,0431
	L	<b>0,0738</b>	<b>0,1</b>	<b>0,1</b>	0	0,0003	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	0,0185	<b>0,1</b>	0,0111	<b>0,0839</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,0621</b>
	CVM	<b>0,1312</b>	<b>0,0551</b>	<b>0,4418</b>	0	0,025	<b>0,4474</b>	<b>0,4996</b>	<b>0,4253</b>	0,0014	<b>0,4559</b>	0,0006	0,0421	<b>0,1612</b>	<b>0,6061</b>	<b>0,7573</b>	<b>0,1303</b>
	W	<b>0,1387</b>	0,0474	<b>0,4098</b>	0	0,0255	<b>0,5214</b>	<b>0,4848</b>	<b>0,4213</b>	0,0012	<b>0,4915</b>	0,0003	0,0317	<b>0,2145</b>	<b>0,599</b>	<b>0,7186</b>	<b>0,1428</b>
	A-D	<b>0,0736</b>	<b>0,0806</b>	<b>0,37</b>	0	0,0429	<b>0,5152</b>	<b>0,3726</b>	<b>0,36</b>	0,0002	<b>0,5919</b>	0,0001	<b>0,0592</b>	<b>0,1056</b>	<b>0,5509</b>	<b>0,5854</b>	<b>0,1547</b>
O-O	J-B	<b>0,0774</b>	0	0	0,0152	<b>0,6789</b>	<b>0,6128</b>	<b>0,6598</b>	0	0	<b>0,0528</b>	0	<b>0,3151</b>	0	<b>0,6781</b>	0,0002	0
	L	0,0054	0,0484	0,0067	0,0118	0,0019	<b>0,1</b>	<b>0,1</b>	0,0058	<b>0,0628</b>	<b>0,1</b>	0,0001	<b>0,1</b>	0,0019	<b>0,1</b>	0,038	0
	CVM	0,007	0,0068	0,0009	0,0256	0,0399	<b>0,2314</b>	<b>0,8654</b>	0,0007	0,0018	0,0453	0	<b>0,182</b>	0,0007	<b>0,4762</b>	<b>0,1048</b>	0
	W	0,0041	0,0039	0,0004	0,0194	0,0303	<b>0,1996</b>	<b>0,8471</b>	0,0003	0,0013	0,0333	0	<b>0,2137</b>	0,0006	<b>0,4468</b>	<b>0,1976</b>	0
	A-D	0,0087	0,005	0,0003	0,0245	<b>0,0557</b>	<b>0,3361</b>	<b>0,8278</b>	0,0004	0,0014	<b>0,0572</b>	0	<b>0,1853</b>	0,0003	<b>0,5247</b>	<b>0,0528</b>	0
O-C	J-B	0,0223	0	0	0	<b>0,8641</b>	0	<b>0,559</b>	0	0	<b>0,8148</b>	0	<b>0,1313</b>	<b>0,1986</b>	<b>0,9285</b>	<b>0,07</b>	0
	L	0,0083	<b>0,0636</b>	0,0116	0	0	0,0168	<b>0,1</b>	0,0122	<b>0,0501</b>	<b>0,1</b>	0	<b>0,0725</b>	<b>0,0518</b>	<b>0,1</b>	<b>0,1</b>	0
	CVM	0,0118	0,0312	0,0045	0	0,0164	<b>0,0577</b>	<b>0,7381</b>	0,0036	0,0022	<b>0,8026</b>	0	<b>0,0949</b>	0,0168	<b>0,8273</b>	<b>0,4197</b>	0
	W	0,0091	0,0231	0,0027	0	0,0119	<b>0,065</b>	<b>0,7058</b>	0,0029	0,0017	<b>0,8096</b>	0	<b>0,106</b>	0,012	<b>0,8054</b>	<b>0,5098</b>	0
	A-D	0,0121	0,0284	0,0016	0	0,0485	<b>0,0964</b>	<b>0,5917</b>	0,0021	0,0009	<b>0,8088</b>	0	<b>0,1083</b>	0,0217	<b>0,9366</b>	<b>0,2962</b>	0
Overnight	J-B	0	0	0	0	0	0,0002	0	0	0	0	0	0	0	0	0	0
	L	0,0011	0	0	0	0	0,0009	0,0062	0	0,0306	0,0007	0,0003	0,0003	0	0	0	0
	CVM	0	0	0	0	0	0,0014	0,0045	0	0,0145	0,0064	0	0,0026	0	0	0	0
	W	0,0001	0	0	0	0	0,01	0,0025	0	0,0103	0,0052	0	0,0014	0	0	0	0
	A-D	0	0	0	0	0	0,0132	0,0011	0	0,0045	0,0009	0	0,0027	0	0	0	0

Źródło: opracowanie własne

Tabela 3. Wyniki testów normalności rozkładów stóp zwrotu dla PKN Orlen w latach 2001-2016

		2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
C-C	J-B	<b>0,054</b>	0	0,0379	0,0305	<b>0,6809</b>	0	<b>0,3235</b>	0	<b>0,0004</b>	<b>0,4511</b>	0	0,0017	<b>0,1178</b>	0	<b>0,8892</b>	0,0011
	L	<b>0,1</b>	0,0068	<b>0,1</b>	0,0141	<b>0,1</b>	<b>0,1</b>	0,0266	<b>0,1</b>	<b>0,1</b>	0,0424	0,0082	<b>0,1</b>	0,029	0,0502	<b>0,1</b>	<b>0,1</b>
	CVM	<b>0,2218</b>	0,0073	<b>0,1946</b>	0,0057	<b>0,4406</b>	<b>0,1257</b>	0,0462	<b>0,1145</b>	<b>0,5973</b>	0,0386	0,0002	<b>0,6745</b>	0,0243	0,0344	<b>0,5472</b>	<b>0,2434</b>
	W	<b>0,2114</b>	0,0074	<b>0,2423</b>	0,0039	<b>0,4482</b>	<b>0,1033</b>	<b>0,0563</b>	<b>0,0959</b>	<b>0,6318</b>	<b>0,0285</b>	<b>0,0001</b>	<b>0,7036</b>	0,0168	0,024	<b>0,5173</b>	<b>0,276</b>
	A-D	<b>0,1571</b>	0,0019	<b>0,1172</b>	0,0063	<b>0,4457</b>	<b>0,092</b>	<b>0,0622</b>	<b>0,9039</b>	<b>0,5637</b>	<b>0,0569</b>	<b>0</b>	<b>0,4508</b>	0,0106	0,0174	<b>0,617</b>	<b>0,21</b>
O-O	J-B	0,0022	0	<b>0,351</b>	<b>0,2688</b>	<b>0,1974</b>	<b>0,2827</b>	<b>0,2617</b>	0	<b>0,1075</b>	<b>0,0722</b>	<b>0</b>	0,0043	<b>0,5579</b>	<b>0,8551</b>	<b>0,8705</b>	0
	L	0,0002	0,0087	0,0421	0,0031	0,0008	<b>0,1</b>	<b>0,1</b>	0,0068	0,0246	<b>0,1</b>	<b>0,0752</b>	0,0298	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	0
	CVM	0,0007	0,0008	<b>0,2216</b>	0,0172	0,0084	<b>0,8539</b>	<b>0,5425</b>	0,0001	<b>0,0896</b>	<b>0,3921</b>	0,0016	0,0331	<b>0,6255</b>	<b>0,8313</b>	<b>0,4751</b>	0
	W	0,0003	0,0005	<b>0,2145</b>	0,0155	0,0083	<b>8437</b>	<b>0,5464</b>	0	<b>0,0769</b>	<b>0,3539</b>	0,0008	0,0245	<b>0,5703</b>	<b>0,8029</b>	<b>0,467</b>	0
	A-D	0,0009	0,0001	<b>0,191</b>	0,0314	0,0186	<b>0,8706</b>	<b>0,4034</b>	0	<b>0,121</b>	<b>0,3658</b>	0,0001	0,0299	<b>0,5341</b>	<b>0,8339</b>	<b>0,6046</b>	0
O-C	J-B	0,0007	0,0002	<b>0,0836</b>	0,0346	<b>0,1799</b>	0,0037	0,128	0	0,0013	0,0265	0	0,0019	0,0029	<b>0,2778</b>	<b>0,8811</b>	0
	L	0,0009	0,0002	0,0049	0,0004	0,0387	<b>0,1</b>	0,0219	0,0072	<b>0,1</b>	<b>0,1</b>	0,0354	0,0274	0,0047	<b>0,1</b>	<b>0,0721</b>	0,0001
	CVM	0,0039	0,0013	<b>0,1141</b>	0,0021	<b>0,1831</b>	<b>0,0619</b>	0,0239	0,0024	<b>0,5558</b>	<b>0,1733</b>	0,0001	0,0463	<b>0,0973</b>	<b>0,1722</b>	<b>0,205</b>	0
	W	0,0003	0,0008	<b>0,1514</b>	0,0014	<b>0,1744</b>	<b>0,0621</b>	0,0275	0,0023	<b>0,511</b>	<b>0,1463</b>	0,0005	0,0036	<b>0,0927</b>	<b>0,1801</b>	<b>0,1861</b>	0
	A-D	0,0055	0,0005	<b>0,104</b>	0,0028	<b>0,2652</b>	0,0388	0,0413	0,0007	<b>0,528</b>	<b>0,1863</b>	0,0003	0,0358	<b>0,0766</b>	<b>0,2009</b>	<b>0,3252</b>	0
Overnight	J-B	<b>0,3247</b>	0	<b>0,1028</b>	0	0	0	0	0	<b>0,0845</b>	0,0006	0	0	0	0	0	0
	L	0	0	0	0	0	0	0	0	0,0297	0,0008	0	0	0,0005	0	0	0
	CVM	0,0003	0	0,0006	0	0	0	0,0001	0	0,0282	0,0012	0	0	0,0013	0	0	0
	W	0,0001	0	0,0003	0	0	0	0	0	0,0219	0,0006	0	0	0,0006	0	0	0
	A-D	0,0008	0	0,0012	0	0	0	0,0001	0	0,0331	0,0007	0	0	0,0014	0	0	0

Źródło: opracowanie własne

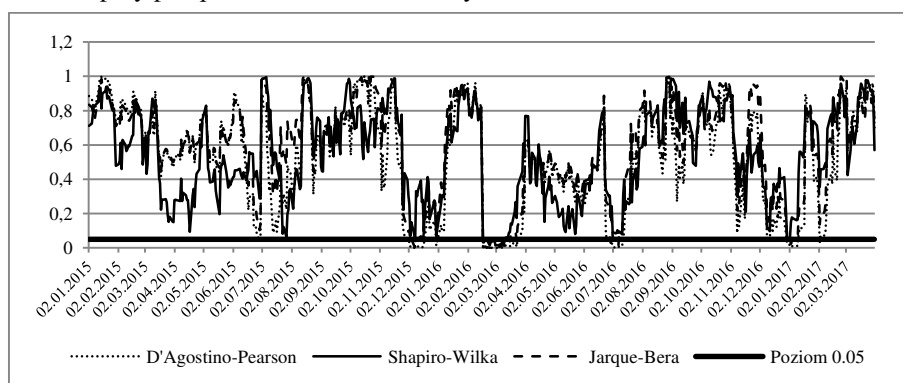
Tabela 4. Liczba rodzajów testów, dla których nie było podstaw do odrzucenia hipotezy zerowej

		2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
KGHM	C-C	0	3	0	0	5	4	3	0	4	5	0	0	0	0	0	4
	O-O	4	0	0	0	5	0	1	0	4	4	0	0	0	0	0	0
	O-C	5	0	1	0	5	0	5	0	0	3	0	4	0	0	0	3
	Overnight	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pekao	C-C	3	5	0	1	4	4	5	0	5	0	2	4	4	5	4	0
	O-O	0	0	0	2	5	5	0	1	3	0	5	0	5	3	0	0
	O-C	1	0	0	1	3	5	0	1	5	0	5	2	5	5	0	0
	Overnight	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PKN Orlen	C-C	5	0	4	0	5	4	3	4	4	2	0	4	1	1	5	4
	O-O	0	0	4	1	1	5	5	0	4	5	1	0	5	5	5	0
	O-C	0	0	4	0	4	3	1	0	4	4	0	0	3	5	5	0
	Overnight	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0

Źródło: opracowanie własne

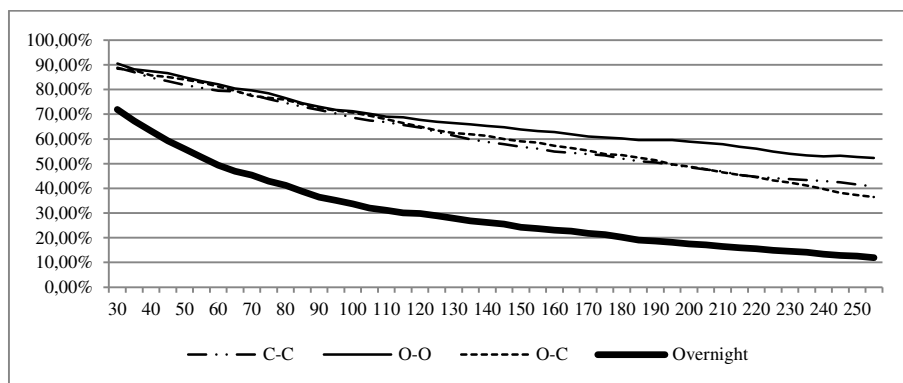
### Testowanie hipotez statystycznych dla K=30, K=100 i K=252 sesje oraz związane z tym badaniem statystyki

Z wykorzystaniem testów: Jarque-Bera, Shapiro-Wilka i D'Agostino-Pearsona zostały przeprowadzone testy dla wszystkich komponentów indeksów WIG20, mWIG40 i sWIG80 dla K=30, 100 i 252 sesji. Dobór K=30 wynika z faktu, że przy liczebności próby ok. 30 elementów przyjmuje się, że rozkład t-Studenta jest zbliżony do rozkładu normalnego. Z kolei K=252 odpowiada w przybliżeniu liczbie sesji w ciągu roku, zaś K=100 jest wielkością pośrednią między K=30 i K=100 i odpowiada liczbie sesji w ciągu ok. 4 miesięcy (przy założeniu 25 sesji w ciągu miesiąca). Ponadto w analizie technicznej średnia o długości 100 sesji traktowana jest jako średnia średnioterminowa. Otrzymane wyniki dla komponentów indeksu WIG20 zostały zamieszczone w tabeli 5, a dla spółki PKN Orlen zamieszczone zostały na rysunku 2 i 3.

Rysunek 2. Wartość parametru  $p$  dla spółki PKN Orlen w okresie 02,01,2015 – 31,03,2017 przy przeprowadzaniu trzech różnych testów i K=30

Źródło: opracowanie własne

Rysunek 3. Odsetek przypadków, przy których nie było podstaw do odrzucenia hipotezy zerowej dla stóp zwrotu cen akcji PKN Orlen i testu Jarque-Bera, w zależności od  $K$  (zmiana  $K$  co 5 jednostek)



Źródło: opracowanie własne

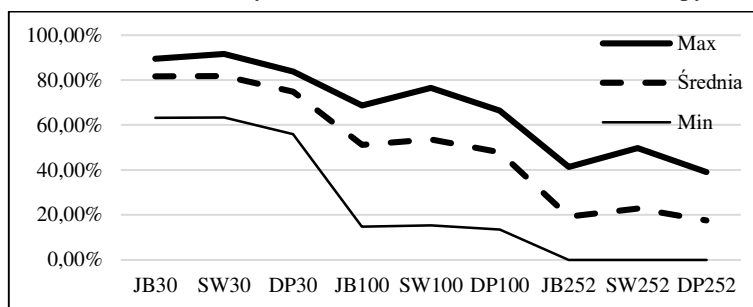
Zwiększanie próby stóp zwrotu w ujęciu *overnight* najszybciej prowadzi do spadku odsetka przypadków, dla których nie było podstaw do odrzucenia hipotezy zerowej. Fakt ten jest zgodny z oczekiwaniami, bowiem jedynie w niewielu przypadkach mają miejsce istotne zdarzenia z otoczenia spółki, zachodzące w czasie kiedy rynek jest zamknięty, powodujące istotne odchylenia ceny otwarcia w stosunku do ostatniej ceny zamknięcia. Znaczny odsetek stóp zwrotu jest w tym przypadku bliski zera. W przypadku stóp zwrotu C-C, O-O oraz O-C, wymagane jest już znacznie poszerzenie horyzontu obserwacji, aby spadł odsetek przypadków, przy których nie było podstaw do odrzucenia hipotezy zerowej. Wykresy analogiczne do rysunku 2 zostały sporządzone także dla innych firm, a na podstawie obserwacji ich przebiegu można wyciągnąć wniosek, że przy rozszerzaniu horyzontu obserwacji, odsetek przypadków, dla których nie było podstaw do odrzucenia hipotezy zerowej jest zbliżony dla stóp zwrotu C-C, O-O i O-C, a zarazem wyższy niż dla stóp zwrotu *Overnight*.

Następnie dla każdej kolumny z tabeli 5 obliczona została: wartość średnia, maksymalna i minimalna. Analogiczne kalkulacje przeprowadzone zostały dla komponentów indeksów mWIG40 i sWIG80. Wyniki dla indeksu WIG20 zamieszczone zostały na rysunkach 4-7 dla wszystkich kalkulowanych stóp procentowych, tj. C-C, O-O, O-C i *Overnight*. Z uwagi na podobieństwo otrzymanych wyników dla indeksów mWIG40 i sWIG80 do rezultatów badań dla indeksu WIG20, dla indeksów mWIG40 i sWIG80 na rysunkach 8 i 9 ukazane zostały wyniki jedynie dla stóp zwrotu C-C.

Na rysunkach 4-13 przyjęto następujące oznaczenie dla wielkości znajdujących się na osi poziomej: dwie pierwsze litery to skrót stosownego testu (JB – Jarque-Bera, SW – Shapiro-Wilka, DP- D'Agostiono-Perasona), zaś

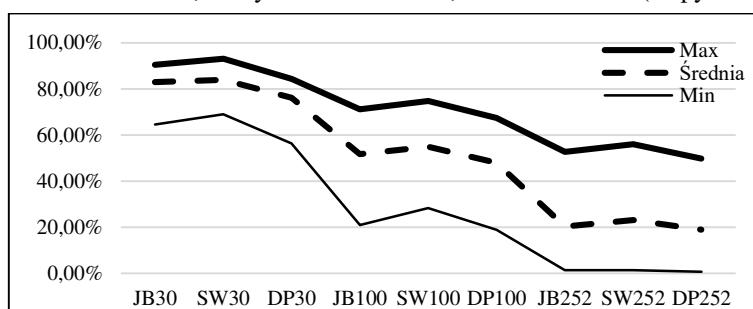
znajdująca się po nich liczba oznacza K (K=30, 100 lub 252). Tak więc skrót np.: JB30 to test Jarque-Bera o K=30 sesji.

Rysunek 4. Wartość średnia, maksymalna i minimalna dla spółek wchodzących w skład indeksu WIG20, różnych testów i K=30, K=100 i K=252 (stopy zwrotu C-C)



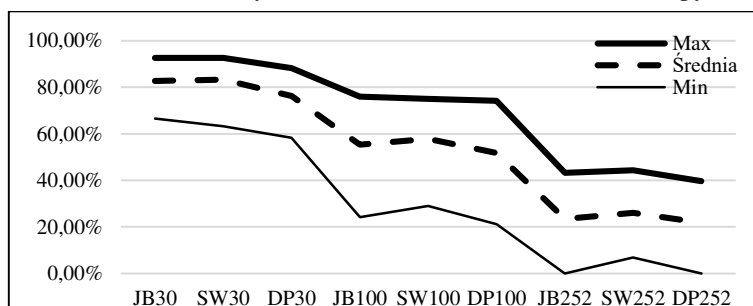
Źródło: opracowanie własne

Rysunek 5. Wartość średnia, maksymalna i minimalna dla spółek wchodzących w skład indeksu WIG20, różnych testów i K=30, K=100 i K=252 (stopy zwrotu O-O)



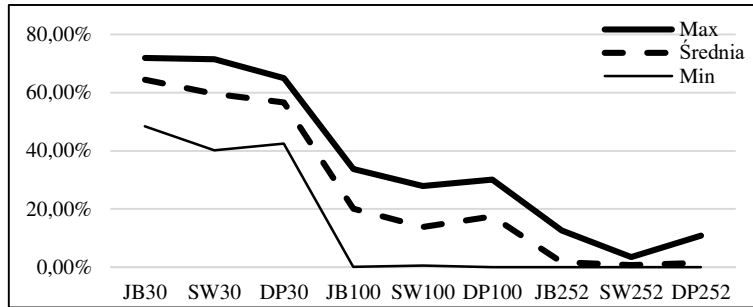
Źródło: opracowanie własne

Rysunek 6. Wartość średnia, maksymalna i minimalna dla spółek wchodzących w skład indeksu WIG20, różnych testów i K=30, K=100 i K=252 (stopy zwrotu O-C)



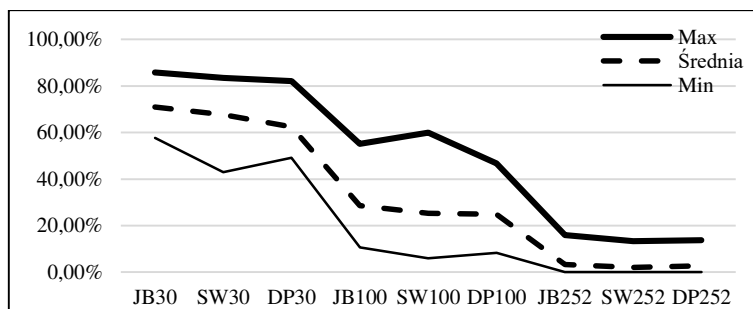
Źródło: opracowanie własne

Rysunek 7. Wartość średnia, maksymalna i minimalna dla spółek wchodzących w skład indeksu WIG20, różnych testów i K=30, K=100 i K=252 sesje (stopy zwrotu *overnight*)



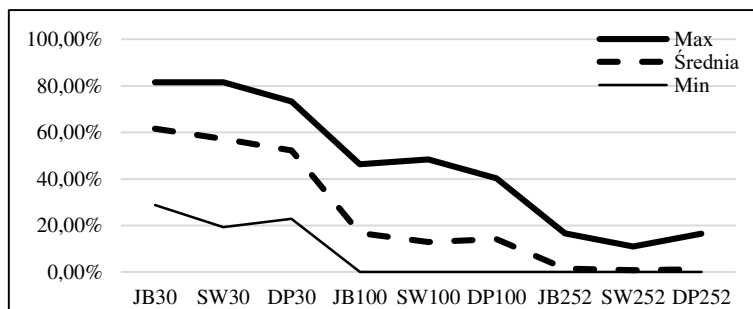
Źródło: opracowanie własne

Rysunek 8. Wartość średnia, maksymalna i minimalna dla spółek wchodzących w skład indeksu mWIG40, różnych testów i K=30, K=100 i K=252 (stopy zwrotu C-C)



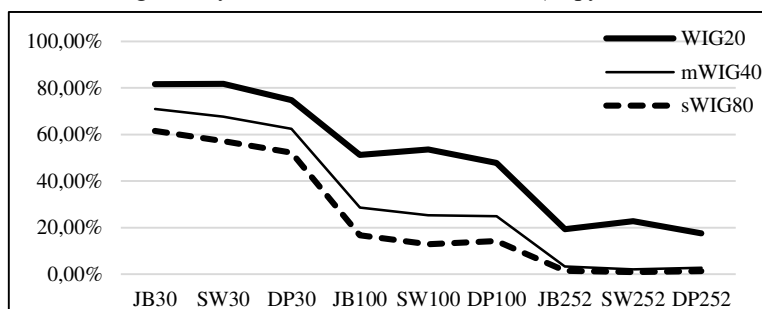
Źródło: opracowanie własne

Rysunek 9. Wartość średnia, maksymalna i minimalna dla spółek wchodzących w skład indeksu sWIG80, różnych testów i K=30, K=100 i K=252 (stopy zwrotu C-C)



Źródło: opracowanie własne

Rysunek 10. Porównanie wartości średnich dla spółek wchodzących w skład poszczególnych indeksów giełdowych dla  $K=30$ ,  $K=100$  i  $K=252$  (stopy zwrotu C-C)

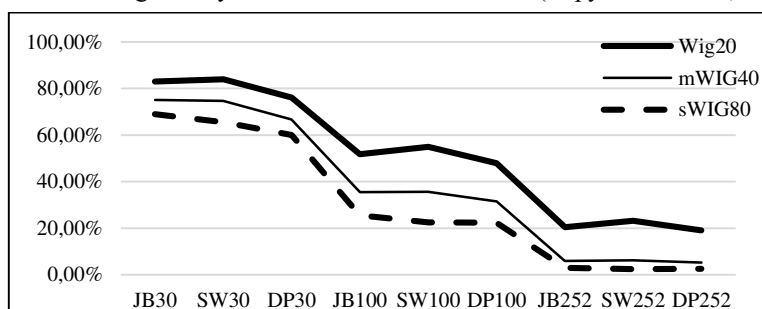


Źródło: opracowanie własne

W przypadku wszystkich analizowanych spółek wydłużenie horyzontu czasowego prowadziło do spadku wartości: średniej, maksymalnej i minimalnej. Fakt ten dowodzi tezy, że rozkład stóp zwrotu spółek notowanych na GPW i wchodzących w skład indeksów WIG20, mWIG40 i sWIG80 jest w krótkich interwałach czasowych rozkładem normalnym, zaś w dłuższych - odbiega od normalnego.

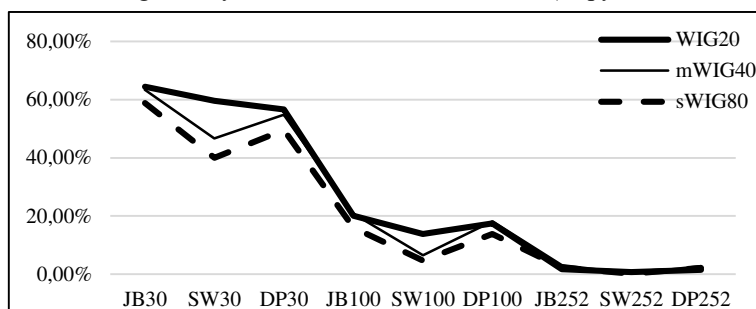
Na rysunkach 10-13 zamieszczone zostało porównanie wartości średnich dla spółek wchodzących w skład poszczególnych indeksów giełdowych. Dla stóp zwrotu C-C i O-O największa wartość średniej została odnotowana w przypadku komponentów indeksu WIG20, wyprzedzając średnie obliczone dla indeksów mWIG40 i sWIG80, Wartości tych średnich malały wraz ze wzrostem liczby sesji, tj. od  $K=30$  poprzez  $K=100$  aż po  $K=252$ . Ponadto różnica między średnimi utrzymywała się na podobnym poziomie, poza przypadkiem kiedy przynajmniej jedna z nich osiągnęła wartość zero. Dla stóp zwrotu O-C i *overnight* wartość średniej była również najwyższa dla spółek z indeksu WIG20, przed komponentami z indeksów mWIG40 i sWIG80, jednak różnice między tymi średnimi wyraźnie zmniejszały się wraz ze wzrostem  $K$  od 30, poprzez 100 aż do 252 sesji.

Rysunek 11. Porównanie wartości średnich dla spółek wchodzących w skład poszczególnych indeksów giełdowych dla  $K=30$ ,  $K=100$  i  $K=252$  (stopy zwrotu O-O)



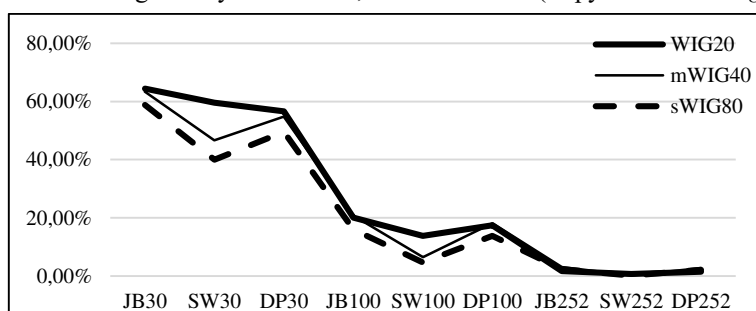
Źródło: opracowanie własne

Rysunek 12. Porównanie wartości średnich dla spółek wchodzących w skład poszczególnych indeksów giełdowych dla K=30, K=100 i K=252 (stopy zwrotu O-C)



Źródło: opracowanie własne

Rysunek 13. Porównanie wartości średnich dla spółek wchodzących w skład poszczególnych indeksów giełdowych dla K=30, K=100 i K=252 (stopy zwrotu overnight)



Źródło: opracowanie własne

Tabela 5. Odsetek przypadków, przy których nie było podstaw do odrzucenia hipotezy zerowej dla K=30, K=100 i K=252 sesje

		Altor	Asseccopol	BZWBK	CCC	Cyfrowy Polsat	Energia	Eurocash	JSW	KGHM	Lotos	LPP	mBank	Orange	Pekao S.A.	PGE	PGNiG	PKN Orlen	PKO BP	PZU	Tauron	
30 sesji	JB	C-C	81	81	82	76	86	63	73	71	90	87	71	83	80	88	89	84	89	89	82	86
		O-O	85	84	82	82	90	65	79	75	85	86	75	89	83	89	87	86	90	86	82	85
		O-C	91	80	82	80	87	68	72	73	87	86	66	83	83	93	91	85	88	89	87	87
	SW	OV	60	70	72	65	68	48	58	52	69	64	62	71	68	69	69	67	72	71	52	64
		C-C	78	81	80	70	87	63	71	73	88	85	70	84	82	90	90	85	91	92	85	87
		O-O	84	86	83	83	93	69	79	75	86	86	76	87	85	89	89	85	91	87	84	84
	DP	O-C	87	83	81	77	91	72	72	73	89	85	63	85	84	93	92	86	92	91	88	86
		OV	49	68	66	50	66	40	48	51	69	61	52	66	66	69	65	64	68	71	49	50
		C-C	73	73	76	67	77	56	65	63	83	81	61	74	74	84	83	76	84	84	77	82
	DP	O-O	79	76	75	74	82	56	68	65	79	81	67	82	77	84	81	80	84	80	76	79
		O-C	83	72	75	72	81	60	64	64	81	81	58	74	77	88	87	78	83	83	83	81
		OV	51	62	63	56	60	43	49	44	60	55	52	63	62	63	61	58	64	65	44	58



		Alior	AssecoPol	BZWBK	CCC	Cyfrowy Polsat	Energa	Eurocash	JSW	KGHM	Lotos	LPP	mBank	Orange	Pekao S.A.	PGE	PGNiG	PKN Orlen	PKO BP	PZU	Tauron	
100 sesji	JB	C-C	50	45	64	38	59	15	29	33	60	60	26	51	48	68	65	56	69	61	53	67
		O-O	58	51	61	41	65	21	37	38	52	67	31	53	57	66	56	62	71	58	44	52
		O-C	72	45	66	41	56	24	32	35	65	64	31	47	58	68	76	62	71	66	63	69
		OV	16	24	30	19	24	0	11	7	27	14	15	25	23	25	22	19	34	30	13	25
	SW	C-C	53	45	63	32	63	15	26	33	65	67	23	55	51	74	68	56	77	69	59	69
		O-O	58	57	61	48	74	28	38	39	56	69	33	57	58	74	61	57	75	59	45	58
		O-C	67	50	65	45	60	31	30	34	69	65	29	55	63	73	75	63	75	71	68	70
		OV	5	18	19	2	17	1	6	3	25	12	3	16	21	28	16	15	28	22	7	7
	DP	C-C	48	42	61	35	54	13	26	30	58	57	23	46	45	65	62	50	65	58	49	66
		O-O	53	47	56	37	61	19	33	35	48	63	29	48	55	61	52	57	67	54	39	49
		O-C	64	39	63	38	53	21	29	34	62	62	28	42	53	64	74	55	66	62	61	65
		OV	15	22	25	16	20	0	10	6	24	12	13	21	21	22	19	16	30	26	11	22
252 sesje	JB	C-C	21	5	29	13	24	0	0	12	21	28	4	7	24	40	22	16	41	25	9	30
		O-O	29	14	22	15	31	2	8	11	8	25	11	28	25	34	17	29	53	22	1	25
		O-C	17	9	34	8	14	0	3	12	25	28	10	18	28	43	38	35	37	32	26	29
		OV	4	2	1	4	0	0	0	0	1	0	0	4	3	1	0	1	13	2	0	1
	SW	C-C	20	3	30	14	34	0	1	15	25	36	5	12	28	50	18	20	47	26	18	31
		O-O	39	20	28	13	41	3	11	16	12	29	12	31	25	41	18	21	56	32	1	25
		O-C	19	10	33	10	20	11	7	13	27	33	8	23	31	41	32	38	44	37	28	32
		OV	0	1	0	0	0	0	0	0	0	0	0	2	3	2	0	0	3	2	0	0
	DP	C-C	18	4	27	11	19	0	0	10	18	25	3	5	23	39	20	15	39	23	9	30
		O-O	25	13	19	14	31	1	7	11	7	22	10	26	23	32	16	26	50	20	1	25
		O-C	16	8	32	7	12	0	3	11	23	27	10	17	25	40	35	34	34	30	25	27
		OV	3	1	0	3	0	0	0	0	1	0	0	4	3	1	0	0	11	2	0	1

Źródło: opracowanie własne

W tabelach 6, 7 i 8 zaprezentowany został ranking spółek wchodzących w skład indeksów WIG20, WIG40 i sWIG80 ze względu na bliskość rozkładu stóp zwrotu tych spółek do rozkładu normalnego. I tak, spośród spółek zaliczanych do indeksu WIG20, dla  $K=30$  sesji i stóp zwrotu C-C, na pierwszym miejscu uplasowały się akcje PKO BP, przed PKN Orlen i Pekao S.A. Na ostatnim miejscu znalazły się akcje Energii. W przypadku komponentów indeksu mWIG40, dla  $K=30$  i stóp zwrotu C-C na pierwszych trzech miejscach sklasyfikowane zostały spółki: Wawel, Amica i GTC, a na ostatnim – Medicalgorithmics. Z kolei, w gronie spółek zaliczonych do indeksu sWIG80, trzy pierwsze pozycje ex aequo zajęły: Agora, Astarta i J.W. Construction, kiedy na ostatnim miejscu znalazły się akcje Biotonu.

Jeśli chodzi o ranking globalny, utworzony na podstawie 12 rankingów cząstkowych, w przypadku poszczególnych indeksów na pierwszych miejscach uplasowały się następujące spółki:

- WIG20: PKN Orlen, Pekao S.A, i PKO BP,
- mWIG40: GTC, Orbis i Bank Handlowy,
- sWIG80: Agora, J.W. Construction i Dębica.

W kwestii indeksu WIG20, pewnym zaskoczeniem może być ostatnia pozycja spółki Energa zwłaszcza, że jest ona notowana na parkiecie GPW od 11.12.2013 roku. Podobne spostrzeżenie można poczynić w przypadku JSW (06.07.2011) czy PZU (12.05.2010), które uplasowały się odpowiednio na 17 i 15 pozycji, Na dalekich pozycjach w rankingu globalnym wśród spółek wchodzących

w skład indeksu mWIG40 sklasyfikowane zostały akcje takich firm jak: GPW (38 pozycja) i Medicalgorithmics (39 pozycja), podczas gdy akcje obu firm kwotowane są na GPW w Warszawie od odpowiednio: 09.11.2010 i 09.11.2011 r. Tymczasem spółki, dla których notowania uwzględnione w badaniu, rozpoczynają się w dniu wprowadzenia systemu WARSET, jak np. Bank Handlowy i Orbis uplasowały się w ścisłej czołówce, tj. na 2 i 3 miejscu. Podobna sytuacja ma miejsce w przypadku komponentów indeksu sWIG80: spółki o dłuższej historii notowań na GPW jak np. Agora, Dębica i Elbudowa, znalazły się na czołowych miejscach globalnego rankingu (1, 3 i 4), podczas gdy firmy, których akcje notowane są w krótkim horyzoncie czasowym: Atal (23.07.2015), Praire (15.09.2015) i Stelmet (25.10.2016) uplasowały się na końcowych pozycjach rankingu (80, 77 i 79). Tego typu wyniki świadczą o tym, że miejsce w rankingu globalnym (a pośrednio w 12 rankingach dla K=30, 100 i 252 sesje oraz różnych stóp zwrotu) nie zależy od daty początkowej notowań akcji spółki na GPW, ale od rozkładu jej stóp zwrotu.

Tabela 6. Ranking spółek wchodzących w skład indeksu WIG20 ze względu na podobieństwo rozkładu stóp zwrotu do rozkładu normalnego

Nazwa spółki	Data pierwszej sesji	30 sesji				100 sesji				252 sesje				Ranking globalny
		C-C	O-O	O-C	O-V	C-C	O-O	O-C	O-V	C-C	O-O	O-C	O-V	
Alior	14.12.2012	15	13	5	16	13	9	5	15	11	4	13	8	13
Assecopol	17.10.2000	15	11	15	6	15	14	15	6	17	13	17	5	14
BZWBK	17.10.2000	12	15	14	4	6	5	8	4	5	11	5	11	6
CCC	02.12.2004	16	16	16	13	16	16	16	13	13	14	18	9	16
Cyfrowy Polsat	06.05.2008	7	2	8	9	9	4	13	9	7	3	14	20	7
Energa	11.12.2013	20	20	19	20	20	20	20	20	20	19	19	20	20
Eurocash	04.02.2005	17	17	18	18	18	18	18	18	19	18	20	20	19
JSW	06.07.2011	18	19	17	17	17	17	17	19	15	15	15	20	17
KGHM	17.10.2000	4	9	7	7	8	13	7	4	9	17	11	10	8
Lotos	09.06.2005	8	7	10	12	7	2	9	15	4	8	7	20	10
LPP	16.05.2001	19	18	20	15	19	19	20	17	18	16	16	20	18
mBank	17.10.2000	11	4	13	5	12	12	14	8	16	5	12	2	11
Orange	17.10.2000	13	13	13	9	14	10	12	7	8	10	9	3	12
Pekao S.A.	17.10.2000	3	3	1	3	2	3	4	5	2	2	1	6	2
PGE	06.11.2009	5	5	2	10	4	9	1	11	10	12	4	20	5
PGNiG	23.09.2005	10	8	11	11	10	6	11	12	12	6	3	8	9
PKN Orlen	17.10.2000	2	1	3	2	1	1	2	1	1	1	2	1	1
PKO BP	10.11.2004	1	7	4	1	5	7	6	2	7	9	6	4	3
PZU	12.05.2010	10	14	7	19	11	15	10	17	15	20	10	20	15
Tauron	30.06.2010	6	11	9		3	11	3	10	3	7	8	12	4

Źródło: opracowanie własne

Tabela 7. Ranking spółek wchodzących w skład indeksu mWIG40 ze względu na podobieństwo rozkładu stóp zwrotu do rozkładu normalnego

Nazwa spółki	Data pierwszej sesji	30 sesji				100 sesji				252 sesje				Ranking globalny
		C-C	O-O	O-C	O-V	C-C	O-O	O-C	O-V	C-C	O-O	O-C	O-V	
Alumetal	2014-07-17	7	37	7	35	5	34	8	36	28	40	32	40	30
Amica	2000-10-17	2	33	11	5	9	35	21	4	40	36	14	2	13
AMRest	2005-04-27	36	19	35	13	39	14	31	25	40	28	25	40	33
Apator	2000-10-17	37	25	34	27	25	25	29	29	12	16	18	23	29
Benefit	2011-04-21	25	12	33	18	31	12	30	23	25	3	40	24	26
Bogdanka	2009-06-25	15	8	5	10	8	15	19	28	9	24	34	40	14
Boryszew	2000-10-17	39	39	40	26	36	38	38	16	19	16	17	9	32
Budimex	2000-10-17	13	28	23	28	12	8	10	25	5	12	9	17	7
CD Projekt	2000-10-17	39	40	38	23	40	40	40	32	29	40	40	40	40
Ciech	2005-02-10	19	7	27	14	18	16	34	26	13	22	40	16	22

Nazwa spółki	Data pierwszej sesji	30 sesji				100 sesji				252 sesje				Ranking globalny
		CC	OO	OC	OV	CC	OO	OC	OV	CC	OO	OC	OV	
Comarch	2000-10-17	24	16	12	29	25	24	16	28	11	20	11	5	15
Echo	2000-10-17	35	22	26	6	27	22	21	2	30	26	24	19	23
Emperia	2002-01-03	32	35	36	35	30	30	33	32	18	22	40	12	34
Enea	2008-11-17	12	9	11	3	11	10	19	1	40	4	11	1	5
Forte	2000-10-17	35	34	32	25	32	30	22	8	26	30	16	8	28
Getin Noble	2012-01-20	8	16	4	23	10	5	6	17	15	9	7	21	6
GPW	2010-11-09	27	24	28	36	29	39	35	35	22	40	40	40	38
Grupa Azoty	2008-06-30	6	11	8	32	14	11	9	34	15	8	5	40	8
GTC	2004-05-06	3	1	3	1	3	5	1	5	1	10	6	4	1
Handlowy	2000-10-17	6	6	3	4	4	9	5	11	2	6	4	6	3
INGBS	2014-07-17	10	13	16	20	16	17	14	30	23	24	22	40	20
Intercars	2004-05-26	16	18	14	8	19	33	28	10	24	11	30	14	17
Kernel	2007-11-23	11	29	23	30	21	19	14	19	6	34	26	40	25
Kety	2000-10-17	10	4	6	24	8	3	2	16	7	1	3	13	4
Kruk	2011-05-10	18	3	19	40	18	3	15	39	18	17	21	40	21
Medicalgorithmics	2011-11-09	40	37	39	39	34	32	36	40	32	31	29	40	39
Millennium	2000-10-17	22	14	17	17	22	23	26	14	10	14	20	25	16
Netia	2000-10-17	30	31	32	15	35	36	37	20	40	35	32	40	37
Neuca	2004-09-30	30	21	20	11	38	26	27	3	40	27	12	7	24
Orbis	2000-10-17	4	2	1	2	2	1	3	8	3	2	1	15	2
Pfleider	2000-10-17	27	5	29	19	26	6	12	21	21	8	13	18	9
PKP Cargo	2013-10-30	17	32	16	37	15	7	7	38	20	5	8	40	18
Polimex	2000-10-17	23	24	25	12	20	22	17	13	8	13	19	11	10
Robyg	2010-11-03	21	10	24	38	37	18	24	38	40	33	33	40	35
Sanok	2000-10-17	32	21	37	17	33	31	32	9	16	19	29	3	27
Stalprodukt	2000-10-17	28	26	30	9	25	28	39	18	28	40	40	21	31
Synthos	2014-07-17	15	17	9	31	13	27	11	22	4	19	16	23	10
Trakcja	2008-04-01	20	27	21	33	29	37	25	33	40	29	27	40	36
Uniwheels	2015-05-08	33	38	18	21	6	13	4	12	31	25	2	40	19
Wawel	2000-10-17	1	30	14	8	1	20	23	6	40	32	24	10	12

Źródło: opracowanie własne

Tabela 8. Ranking spółek wchodzących w skład indeksu sWIG80 ze względu na podobieństwo rozkładu stóp zwrotu do rozkładu normalnego

Nazwa spółki	Data pierwszej sesji	30 sesji				100 sesji				252 sesje				Ranking globalny
		CC	OO	OC	OV	CC	OO	OC	OV	CC	OO	OC	OV	
11BIT	2010-10-28	33	58	25	54	56	60	55	71	80	25	80	80	53
ABC Data	2010-06-17	43	53	10	57	51	37	33	61	80	80	30	80	47
AB PL	2006-09-21	7	13	41	20	7	19	17	13	24	23	80	29	15
Autogaz	2011-08-11	45	15	74	70	67	23	72	54	80	46	80	80	58
Agora	2000-10-17	3	5	1	7	1	3	1	11	2	2	1	2	1
Alchemia	2000-10-17	73	70	61	36	68	69	70	63	80	80	80	80	72
Altus TFI	2014-08-20	43	36	12	64	65	53	7	80	80	80	80	80	54
Arctic	2009-10-23	24	28	22	45	60	60	63	64	80	80	80	80	55
Asbis	2007-10-30	27	28	22	25	22	16	13	25	80	19	14	33	17
Assecobs	2007-11-19	12	38	66	55	13	11	33	57	31	38	36	80	37
Assecosee	2009-10-28	56	68	77	57	43	28	76	69	80	39	80	80	63
Astarta	2006-08-17	3	9	3	5	20	49	31	28	80	80	80	80	36
Atal	2015-07-23	79	76	80	77	67	80	49	80	80	80	80	80	80
Autopartner	2016-06-06	72	72	67	28	80	75	70	80	80	80	80	80	75
Bioton	2005-03-16	80	80	74	77	8	11	5	4	27	34	36	23	35
BOS	2000-10-17	17	43	76	39	12	34	52	18	10	22	80	29	33
Braster	2012-12-20	69	72	66	41	73	73	75	65	80	80	80	80	74
Bumech	2009-01-14	33	43	46	68	41	45	46	49	80	80	36	80	52
Bytom	2000-10-17	70	66	66	31	45	62	60	2	30	80	80	23	47
Cigames	2007-11-30	48	48	44	46	65	30	50	22	80	14	80	80	45
CLNPHARMA	2016-12-01	65	75	71	45	80	28	80	7	80	80	80	80	67
Comp	2005-01-14	24	8	54	28	24	14	36	16	80	35	80	14	27
Cormay	2008-08-20	45	62	50	9	47	58	46	45	80	44	31	80	44
CPGroup	2014-03-07	78	74	78	74	73	64	57	32	80	80	80	23	68
Dębica	2000-10-17	25	16	7	13	28	14	6	15	12	11	6	4	3
Dom Development	2006-10-24	9	12	10	66	5	8	13	52	20	24	12	23	10
EkoExport	2000-10-17	65	67	37	11	62	69	65	44	80	80	80	80	61
Elbudowa	2000-10-17	9	4	31	21	10	25	19	11	8	11	16	9	4
Elemental	2012-07-12	6	25	17	5	2	7	4	48	7	8	13	80	7

Nazwa spółki	Data pierwszej sesji	30 sesji				100 sesji				252 sesje				Ranking globalny
		C-C	O-O	C-O	O-V	C-C	O-O	C-O	O-V	C-C	O-O	C-O	O-V	
Enter	2016-01-18	26	73	43	61	80	80	59	80	80	80	80	80	70
Famur	2006-08-04	29	18	8	13	19	39	17	22	10	8	8	23	5
Ferro	2010-04-14	22	2	55	39	5	1	40	24	18	43	80	80	25
Getin	2001-05-10	29	29	18	18	35	55	51	34	16	19	19	80	23
Groclin	2000-10-17	74	63	68	43	80	66	20	4	80	80	36	8	49
Idea Bank	2015-04-29	53	70	60	61	30	72	61	70	80	80	80	80	69
Impexmetal	2000-10-17	37	18	6	11	58	29	44	8	80	36	80	11	28
Izoblok	2011-12-22	71	36	74	64	34	6	23	32	80	42	80	80	49
J. W. Construction	2007-06-04	3	22	3	9	3	14	4	7	4	16	4	8	2
Kogeneracja	2000-10-17	46	45	50	61	39	39	48	26	14	31	39	6	34
Kopex	2000-10-17	43	46	29	24	28	20	17	39	20	14	17	80	22
LCCorp	2007-06-29	20	20	15	35	30	43	27	7	80	34	29	29	21
Lentex	2000-10-17	17	22	17	17	31	34	27	28	27	29	12	6	13
Livechat	2014-04-11	55	66	53	71	17	47	40	68	1	37	27	80	41
Mabion	2010-08-10	50	32	25	7	60	22	10	9	80	4	26	80	24
Mangata	2005-03-04	31	31	47	53	33	35	17	57	17	46	15	29	26
MCI	2001-02-01	19	19	6	16	16	44	13	22	29	31	19	23	12
Mennica	2000-10-17	14	11	37	23	16	16	29	18	80	41	36	33	19
MNLPGGroup	2013-10-28	78	79	80	79	65	71	80	48	80	80	80	80	78
Monnari	2006-12-20	57	58	41	28	62	71	67	38	80	80	80	80	62
Mostostal Zabrze	2000-10-17	43	43	22	31	21	51	23	30	15	32	23	4	18
Newag	2007-06-29	52	43	41	61	49	65	71	61	80	80	80	80	66
Opono	2007-09-12	19	8	22	15	54	37	41	41	80	9	25	80	32
PBG	2004-07-02	38	23	12	20	39	31	25	22	3	5	3	35	11
PCC Rokita	2014-06-25	63	62	27	78	52	41	4	80	80	80	39	80	55
PCM	2014-04-16	68	53	59	72	58	52	66	80	80	80	80	80	71
Pelion	2000-10-17	4	6	6	41	6	4	9	61	7	3	8	80	9
PEP	2005-05-13	60	43	38	50	56	47	28	54	80	80	9	23	42
PHN	2013-02-13	12	33	27	67	49	17	19	55	80	41	80	80	40
Police	2005-07-14	39	50	53	32	43	49	57	58	12	17	22	80	39
Polnord	2000-10-17	22	24	43	1	24	41	42	14	80	27	28	13	20
Polwax	2014-10-06	61	49	50	69	80	75	68	66	80	80	80	80	73
Praire	2015-09-15	76	77	76	80	73	80	80	35	80	80	80	80	77
Quercus	2008-09-11	10	30	37	39	37	18	43	33	80	8	6	80	29
Rafako	2000-10-17	48	51	33	14	37	58	40	37	22	27	21	35	30
Rainbow	2007-10-09	14	15	28	22	25	26	23	51	5	29	21	33	14
RankProgress	2010-07-08	37	38	34	31	45	50	35	51	80	14	80	10	38
Selvita	2008-04-18	50	3	57	3	19	2	40	2	27	2	12	1	6
Snieżka	2003-12-31	30	10	66	50	14	5	55	41	29	20	24	80	31
Stalexport	2000-10-17	55	62	56	52	69	58	63	30	80	80	80	23	59
Stelmet	2016-10-25	76	79	70	64	80	80	74	63	80	80	80	80	79
TIM	2000-10-17	63	36	58	35	46	54	59	44	23	80	80	29	45
Torpol	2014-09-05	60	26	13	73	70	67	48	80	80	80	80	80	65
Ursus	2007-12-27	68	58	45	51	50	64	65	42	80	80	80	33	60
Vistula	2000-10-17	34	47	23	35	10	22	25	13	21	22	39	23	16
Wielton	2007-11-28	15	44	14	3	28	25	9	48	14	16	2	13	8
Wirtualna Polska	2015-05-07	35	1	31	75	33	9	55	80	80	80	80	80	51
Workservice	2012-04-26	60	58	53	66	54	42	35	68	80	80	80	80	64
XTB	2016-05-17	68	54	66	47	80	80	80	80	80	80	80	80	76
ZEPAK	2012-10-30	5	64	33	50	12	34	30	36	80	80	80	80	43
ZPUE	2000-10-17	52	59	70	43	41	62	73	24	80	80	80	29	57

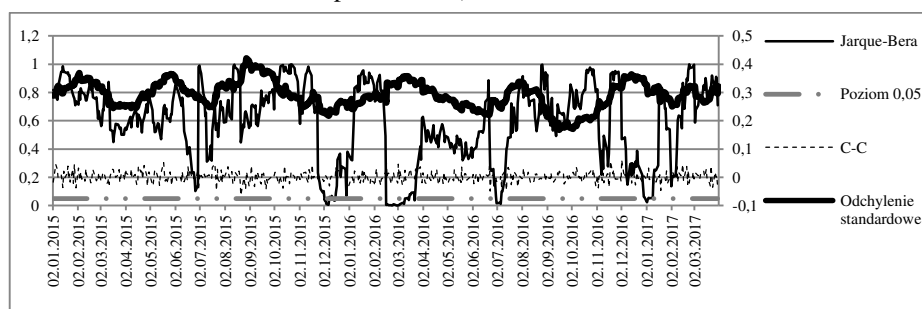
Źródło: opracowanie własne

## ZAKOŃCZENIE

Część przeprowadzonych w artykule obliczeń jednoznacznie dowodzi tezy, że rozkład stóp zwrotu cen akcji notowanych na GPW w Warszawie nie jest rozkładem normalnym, potwierdzając tym samym wyniki uzyskane przez innych badaczy [Bołt, Miłobędzki 1994], [Rokita 2000], [Osińska 2006], [Witkowska, Kompa 2007]. Uwaga ta dotyczy stóp zwrotu w ujęciu C-C, W artykule wykazano także, że rozkład stóp zwrotu w innych trzech ujęciach O-O, C-O i *Overnight* nie jest rozkładem normalnym dla stop zwrotu cen akcji – komponentów indeksów:

WIG20, mWIG40 i sWIG80. Ponadto udowodniona została teza, że rozkład stóp zwrotu może być rozkładem normalnym jedynie w określonych interwałach czasowych – uzyskane rezultaty pokrywają się w części z wynikami pracy Piasecki i Tomasik [Piasecki, Tomasik 2013], którzy wykazali normalność rozkładu stóp zwrotu w określonych falach wzrostowych i spadkowych cen akcji na GPW w Warszawie.

Rysunek 14. Zachowanie parametru  $p$  dla spółki PKN Orlen,  $K=30$ , test Jarque-Bera, stopa zwrotu C-C oraz zannualizowane odchylenie standardowe stóp zwrotu (wartości parametru  $p$  i poziom 0,05 – lewa skala, stopa zwrotu C-C i odchylenie standardowe - prawa skala)



Źródło: opracowanie własne

Z wykorzystaniem parametru  $p$  stworzony został również ranking spółek wchodzących w skład poszczególnych indeksów giełdowych, ze względu na możliwość przybliżenia rozkładu ich stóp zwrotu rozkładem normalnym w horyzoncie  $K=30$ ,  $K=100$  i  $K=252$  sesje. W wyniku czego otrzymano wniosek, że pozycja spółki w rankingu nie jest zależna do daty jej debiutu giełdowego, a tym samym liczby możliwych do obliczenia dla niej stóp zwrotu, ale od rozkładu tychże stóp zwrotu.

Analiza wyników otrzymanych dla  $K=30$  sesji pozwala wyciągnąć wniosek, że w przypadku tak krótkich interwałów czasowych, gwałtowna zmiana ceny prowadzi do silnego spadku wartości parametru  $p$ . Proces ten zostały zilustrowany na rysunku 14, na którym zamieszczono: zmiany ceny akcji PKN Orlen w ujęciu C-C, parametr  $p$  i zannualizowane odchylenie standardowe stóp zwrotu. I tak np. w okresie małej zmienności cen w czerwcu 2016 r., zwyżka ceny w dniu 23 (sesja poprzedzająca referendum w Wielkiej Brytanii dotyczące wyjścia tego kraju z UE) doprowadziła do spadku wartości parametru  $p$  poniżej 0,05 (0,0401). Wyjaśnienie spadku wartości parametru  $p$  poniżej 0,05 dla  $K=100$  i  $K=252$  sesje jest zagadnieniem bardziej złożonym i wymaga dalszych badań. Podobne badania powinny być prowadzone również dla aktywów na rynku surowców (*commodities*) i FX.

## BIBLIOGRAFIA

- Bachelier L. (1900) Theorie de la speculation. *Annales de l'Ecole Normal Supérieure*, XVII (3), 21-86.
- Barunik J., Vacha I., Vosvrda M. (2010) The Behavior of Central European Stock Markets During the Financial Crisis. *Czech Economic Review*, 3, 281-294.
- Bolt T., Miłobędzki P. (1994) The Warsaw Stock Exchange in the Period 1991-1993. *Quantitative Problems of Return, Economics of Planning*, 27, 211-226.
- Fiszeder P. (2000) Statystyczne i dynamiczne własności stóp zwrotu na przykładzie światowych indeksów giełdowych. *Nasz Rynek Kapitałowy*, 109, 187-197.
- Ghahfarokhi M., Ghahfarokhi P. (2009) Applications of Stable Distributions in Time Series Analysis, Computer Sciences and Financial Markets. *International Scholarly and Scientific Research & Innovation*, 3 (49), 132-136.
- Kendall M. (1953) The Analysis of Economic Time Series – Part I: Prices. *Journal of Royal Statistical Society*, 116, Series A, 11-25.
- Osborne M. (1959) Brownian Motion in the Stock Market. *Operations Research*, 7 (2), 145-173.
- Osińska M. (2006) *Ekonometria finansowa*. PWE, Warszawa,
- Piasecki K., Tomasik E. (2013) *Rozkłady stóp zwrotu z instrumentów polskiego rynku kapitałowego*. edu-Libri, Kraków-Warszawa.
- Scalas E., Kim K. (2007) The art of fitting financial time series with Levy stable distributions. *Korean Journal of Physics*, 50 (1), 105-111.
- Witkowska D., Kompa K. (2007) Analiza własności stop zwrotu akcji wybranych spółek. [w:] Tarczyński W. (red.) *Rynek kapitałowy, skuteczne inwestowanie. Materiały konferencyjne, część I*. Wydawnictwo Naukowe Uniwersytetu w Szczecinie.

**NORMAL DISTRIBUTION OF RETURNS OF COMPONENTS  
OF THE FOLLOWING WSE INDEXES: WIG20, MWIG40 AND  
SWIG80**

**Abstract:** The article verified the hypothesis regarding normal distribution of returns of shares - components of the following Warsaw Stock Exchange indexes: WIG20, mWIG40 and sWIG80 in the period from the first quotation of the given share (however not earlier than the implementation of the WARSET transaction system on the WSE, i.e. 17.10.2017) until 31.03.2017. The analyzed rates of return were calculated in the following outline: closing-closing, opening-opening, opening-closing and overnight. The companies were ranked due to the convergence of their return to normal distribution with the use of the following tests: Jarque-Bera, Shapiro-Wilk and D'Agostino-Pearson.

**Keywords:** normal distribution, return rates, stock indices, ranking of companies

## ZRÓŻNICOWANIE I DYNAMIKA ZMIAN POWIERZCHNI UŻYTKÓW ROLNYCH WEDŁUG GRUP OBSZAROWYCH GOSPODARSTW W KRAJACH UNII EUROPEJSKIEJ W OKRESIE 2010-2013

**Jadwiga Bożek, Karol Kukuła, Czesław Nowak**

Wydział Rolniczo-Ekonomiczny

Uniwersytet Rolniczy w Krakowie

e-mail: rrbozek@cyf-kr.edu.pl; ksm@ur.krakow.pl; nowak@ar.krakow.pl

**Streszczenie:** W pracy przedstawiono zróżnicowanie i dynamikę zmian powierzchni użytków rolnych skupionych w gospodarstwach o różnej wielkości w krajach Unii Europejskiej w latach 2010 – 2013. Badania przeprowadzono na podstawie danych Eurostat-u. Uwzględniono następujące grupy gospodarstw: do 5 ha UR, 5-20, 20-50, 50 i więcej ha. Posługując się wybranymi metodami statystycznej analizy struktur, określono skalę i kierunek zmian badanej struktury w poszczególnych krajach. W oparciu o klasyfikację rozmytą wyodrębniono cztery grupy krajów o podobnym odsetku UR skupionych w gospodarstwach o różnej wielkości.

**Słowa kluczowe:** Słowa kluczowe: powierzchnia użytków rolnych, Unia Europejska, dynamika, klasyfikacja rozmyta

### WSTĘP

Struktura agrarna krajów Unii Europejskiej jest bardzo zróżnicowana. W 2010 roku w Polsce ponad 55% ogólnej liczby gospodarstw stanowiły gospodarstwa najmniejsze, do 5 ha. Gospodarstwa te użytkowały 13,9% ogółu użytków rolnych. Podobna sytuacja występowała także na Litwie, w Słowenii i Rumunii. W krajach charakteryzujących się najlepszą strukturą gospodarstw, takich jak Dania, Niemcy, Wielka Brytania udział gospodarstw najmniejszych nie przekracza 10% ogólnej ich liczby, a skupiają one nie więcej niż 1% powierzchni wszystkich użytków rolnych w danym kraju. Jednocześnie w krajach tych gospodarstwa największe, o powierzchni co najmniej 50 ha, skupiają około 80% użytków rolnych, podczas gdy w Polsce odsetek ten wynosi 30%. Od przystąpienia

do UE w nowych krajach członkowskich następują znaczne przemiany w strukturze obszarowej gospodarstw rolnych, spowodowane głównie przesłankami ekonomicznymi, ale także działaniami realizowanymi na poziomie polityki poszczególnych państw oraz mechanizmami wprowadzanymi w ramach Wspólnej Polityki Rolnej UE [Babiak 2010, Klepacki, Żak 2013, Poczta 2013, Stańko, Mikuła 2016]. Przemiany te są uważnie śledzone, zwłaszcza w kontekście WPR. Niniejsza praca wpisuje się w tę tematykę, przedstawiając przemiany, jakie nastąpiły w strukturze obszarowej gospodarstw rolnych w krajach UE w latach 2010-2013.

Strukturę obszarową gospodarstw rolnych można rozpatrywać w dwóch aspektach: pod względem liczby gospodarstw według grup wielkościowych oraz pod względem zajmowanej przez nie powierzchni użytków rolnych. W pracy rozpatrywany jest drugi aspekt, tj. powierzchnia użytków rolnych zajmowanych przez gospodarstwa z poszczególnych grup obszarowych. Celem pracy jest określenie skali, kierunków przemian oraz stopnia zróżnicowania powierzchni UR skupionych w gospodarstwach różnej wielkości w krajach UE w latach 2010–2013. Analizę przeprowadzono na podstawie danych Eurostat-u przedstawiających wyniki dwóch ostatnich badań gospodarstw rolnych w krajach UE w 2010 i 2013 roku. Dane dotyczące roku 2010 są wynikami Powszechnego Spisu Rolnego, natomiast 2013 roku - wynikami badania reprezentacyjnego. Uwzględniono następujące grupy obszarowe gospodarstw: do 5 ha użytków rolnych, 5-20 ha, 20-50 ha, 50 i więcej ha. Posługując się wybranymi metodami statystycznej analizy struktur, określono stopień i kierunek zmian badanej struktury w poszczególnych krajach. W oparciu o klasyfikację rozmytą wyodrębniono grupy krajów charakteryzujących się podobnym odsetkiem powierzchni użytków rolnych skupionych w gospodarstwach z poszczególnych grup obszarowych w roku 2010 i 2013, otrzymując w ten sposób syntetyczny obraz zróżnicowania i przemian badanej struktury w krajach UE.

## METODA BADAWCZA

W pracy zastosowano wybrane metody statystycznej analizy struktur. W celu oceny stopnia zmian, jakie zaszły w strukturze w pewnym okresie czasu zastosowano miernik określony następująco [Kukuła 1989]:

jeżeli  $\alpha$  jest strukturą udziałową badaną w czasie  $t = 0, 1, \dots, n$ , złożoną z  $r$  składowych, tzn. dana jest macierz  $[\alpha_{tk}]_{(t=0,1,\dots,n; k=1,\dots,r)}$ , gdzie

$$\sum_{i=1}^r \alpha_{ti} = 1 \quad \text{oraz} \quad 0 \leq \alpha_{ti} \leq 1 \quad t = 0, 1, \dots, n, \quad k = 1, 2, \dots, r$$



$$Wtedy \quad v_{t,t-\tau} = \frac{\sum_{i=1}^r |\alpha_{it} - \alpha_{(t-\tau)i}|}{2}$$

określa stopień zmian struktury w okresie od  $t - \tau$  do  $t$ . Miernik ten przybiera wartości z przedziału  $[0;1]$ . Wysoka jego wartość (tzn. bliska 1) wskazuje, że w strukturze zaszły duże zmiany w okresie od  $t - \tau$  do  $t$ .

Do grupowania krajów pod względem podobieństwa struktury agrarnej zastosowano klasyfikację rozmytą [Zadeh 1965]. Metoda ta umożliwia klasyfikację zbioru obiektów wielowymiarowych  $P_1, P_2, \dots, P_n$ , (w tym przypadku krajów) opisanych przez wartości  $r$  zmiennych:  $X_1, X_2, \dots, X_n$ , (w pracy  $X_i$  oznacza udział powierzchni UR skupionej w gospodarstwach z 1-tej grupy wielkościowej w ogólnej powierzchni UR w danym kraju). Klasyfikacja rozmyta polega na wyznaczeniu dla każdego obiektu  $P_i$  stopnia jego przynależności (podobieństwa) do klasy  $S_j$  czyli tzw. funkcji przynależności  $f_{S_j}(P_i)$ . Funkcja ta przyjmuje wartości z przedziału  $[0,1]$ , przy czym spełniony jest warunek:

$$\sum_{j=1}^K f_{S_j}(P_i) = 1 \quad (i = 1, \dots, n).$$

Wyższa wartość  $f_{S_j}(P_i)$  oznacza większe podobieństwo obiektu  $P_i$  do klasy  $S_j$ .

Obiekty, dla których stopnie przynależności do tej samej klasy są duże – są bardzo podobne, natomiast obiekty, dla których stopnie przynależności do różnych klas są duże – są mało podobne.

Istnieje kilka metod tworzenia klasyfikacji rozmytej [Jajuga 1984]. W pracy została zastosowana metoda iteracyjna, wykorzystująca pojęcie rozmytego środka ciężkości [Jajuga 1984, Bożek 2013]. Otrzymaną w ten sposób klasyfikację rozmytą przekształcono następnie w klasyfikację klasyczną, przyjmując, że obiekt  $P_i$  należy do klasy (grupy typologicznej)  $S_j$ , gdy

$$f_{S_j}(P_i) = \max_l f_{S_l}(P_i).$$

Przedstawiona metoda klasyfikacji obiektów wielowymiarowych jest przydatna w badaniach przestrzenno-czasowych struktury agrarnej [Bożek, Bożek 2011].

## WYNIKI BADAŃ

W latach 2010-2013 w krajach Unii Europejskiej nastąpiły zmiany w powierzchni UR, co obrazuje tabela 1 (w tabeli tej, dla łatwiejszego uchwycenia tendencji zmian, zgrupowano osobno kraje UE-15 i UE-12). Z powodu braku wiarygodnych danych dla Grecji za 2010 rok, kraj ten został pominięty w analizie

dotyczącej tego roku<sup>1</sup>. W badaniach pominięto również Malte i Luksemburg z powodu bardzo małej liczebności gospodarstw, znacznie odbiegającej od pozostałych krajów.

W większości krajów UE-15 powierzchnia użytków rolnych zmniejszyła się w granicach 1-2% (tabela 1). W ujęciu bezwzględny największe ubytki odnotowano we Włoszech, Hiszpanii i Austrii. Natomiast w Wielkiej Brytanii nastąpił wzrost powierzchni UR. W nowych państwach członkowskich przemiany następowały w różnych kierunkach. W siedmiu spośród państw UE-12 powierzchnia UR zwiększyła się, przy czym największe przyrosty odnotowano w Bułgarii (o 175,5 tys. ha), na Litwie (118,7 tys. ha) i Łotwie (81,4 tys. ha). W pozostałych krajach UE-12 powierzchnia UR zmniejszyła się, przy czym najbardziej w Rumunii (o 250,2 tys. ha).

W ujęciu procentowym największe zmiany nastąpiły w powierzchni UR zajmowanej przez gospodarstwa bardzo małe, do 5 ha użytków rolnych, jednakże w ujęciu bezwzględny gospodarstwa te zajmują ogólnie niewielkie powierzchnie UR (tabela 2).

We wszystkich krajach UE-15, oprócz Irlandii i Wielkiej Brytanii powierzchnia UR skupiona w gospodarstwach 5-20 ha spadła. Natomiast w nowych państwach członkowskich odnotowano przemiany w różnych kierunkach: w sześciu krajach powierzchnia ta wzrosła, a w pozostałych spadła. Średni spadek (dla wszystkich państw, w których nastąpiło zmniejszenie powierzchni UR w gospodarstwach 5-20 ha) wynosił 6,8%.

Powierzchnia UR skupiona w gospodarstwach 20-50 ha zmniejszyła się we wszystkich krajach UE-15, z wyjątkiem Portugalii. Natomiast w ośmiu nowych krajach członkowskich gospodarstwa 20-50 ha w 2013 roku użytkują większą powierzchnię, niż w 2010 roku, przy czym największy przyrost procentowy miał miejsce w Bułgarii (+11%) i Polsce (+9%). W pozostałych trzech krajach powierzchnia ta nieco się zmniejszyła. Średni spadek (dla wszystkich państw, w których nastąpiło zmniejszenie powierzchni UR w tej grupie gospodarstw) wynosił -5,4%, natomiast średni wzrost (dla państw, w których nastąpiło zwiększenie powierzchni UR) wynosił 6,5%.

Gospodarstwa największe w większości krajów zwiększyły powierzchnię użytkowanej ziemi średnio o 4,6%. W 10 krajach odnotowano spadek średnio o 2,2%.

Chociaż przeprowadzone badania obejmują krótki okres czasu (2010-2013), to w badanej strukturze nastąpiły widoczne zmiany (tabela 2). Największe zmiany miały miejsce w strukturze Finlandii, dla której stopień zmian strukturalnych wynosi 0,057, Łotwy (0,056) i Litwy (0,054), a także Polski (0,031), Włoch (0,030), Holandii (0,028). Najmniejsze zmiany (poniżej 0,006) odnotowano w Hiszpanii, Irlandii, Belgii, Wielkiej Brytanii.

---

<sup>1</sup> Powierzchnia UR ogółem w Grecji w 2010 r. wynosiła: 3478 tys. ha wg Agricultural Census 2010 – main results - utworzony w 2013 roku i 5177,5 tys. ha wg Eurostat File: Agricultural Holdings, 2000-10, YB14.png - utworzony 4 June 2014.

Tabela 1. Zmiana powierzchni użytków rolnych według grup obszarowych gospodarstw i krajów Unii Europejskiej w okresie 2010- 2013

Kraj	ogółem		Grupy obszarowe gospodarstw w ha UR			
			do 5	5-20	20-50	>=50
	tys. ha	2010=100	2010=100			
Austria	-151,3	94,7	91,4	92,6	97,1	94,3
Belgia	-50,1	96,3	73,4	99,8	96,7	96,1
Dania	-27,6	99,0	86,8	91,0	89,8	100,9
Finlandia	-33,4	98,5	51,7	81,7	87,7	108,3
Francja	-97,9	99,6	87,3	88,5	88,3	101,7
Hiszpania	-452,5	98,1	98,5	99,5	97,3	98,0
Holandia	-24,8	98,7	90,1	92,6	93,5	104,1
Irlandia	-32,0	99,4	101,7	101,3	99,0	99,1
Niemcy	-4,4	100,0	82,6	94,0	93,8	101,9
Portugalia	-26,6	99,3	85,4	96,4	111,1	100,4
Szwecja	-37,7	98,8	87,2	97,1	91,6	100,7
Wielka Brytania	214,5	101,3	103,8	100,0	96,1	101,8
Włochy	-757,2	94,1	74,6	94,4	97,1	99,2
Bułgaria	175,5	103,9	78,4	100,7	111,0	105,3
Cypr	-9,1	92,3	92,9	87,8	91,0	95,6
Czechy	8,0	100,2	143,1	116,2	106,7	99,5
Estonia	16,6	101,8	95,3	97,4	99,3	102,6
Litwa	118,7	104,3	80,6	95,8	96,0	114,9
Łotwa	81,4	104,5	95,6	83,2	100,2	114,7
Polska	-37,4	99,7	93,8	93,8	109,0	104,0
Rumunia	-250,2	98,1	94,4	109,4	104,6	96,9
Słowacja	6,1	100,3	95,8	127,3	106,3	99,6
Słowenia	3,1	100,6	97,6	98,5	106,1	106,3
Węgry	-29,8	99,4	88,7	101,5	103,4	99,4

Źródło: obliczenia własne na podstawie [www.europa.eu/eurostat](http://www.europa.eu/eurostat) (data utworzenia: październik 2013, grudzień 2015)

Najbardziej zmieniły się udziały w UR gospodarstw największych, o powierzchni co najmniej 50 ha. W większości krajów był to wzrost udziałów, a średnia zmiana wynosiła 2,33 punkty procentowe. Najmniejsze zmiany nastąpiły w udziałach powierzchni zajmowanej przez gospodarstwa najmniejsze, do 5 ha (średnio 0,96 pp.) i był to przeważnie spadek udziałów.

Tabela 2. Powierzchnia użytków rolnych (w %) w gospodarstwach rolnych według grup obszarowych i krajów UE w 2010 i 2013 roku

Kraj	2010				2013				V <sub>2013,2010</sub>
	Grupy obszarowe gospodarstw w ha UR				Grupy obszarowe gospodarstw w ha UR				
	do 5	5-20	20-50	>=50	do 5	5-20	20-50	>=50	
	w %				w %				
UE (28)	6,8	12,6	14,2	66,3	6,9	12,7	14,2	66,0	
Austria	4,1	23,1	34,7	38,1	4,0	22,6	35,5	37,9	0,009
Belgia	1,4	10,1	29,5	59,0	1,1	10,4	29,6	58,9	0,005
Dania	0,1	6,4	11,2	82,2	0,1	5,9	10,2	83,8	0,016
Finlandia	0,8	11,2	30,8	57,2	0,4	9,3	27,4	62,9	0,057
Francja	1,0	3,8	10,8	84,5	0,8	3,4	9,6	86,2	0,018
Hiszpania	4,4	10,8	14,3	70,6	4,4	10,9	14,2	70,6	0,002
Holandia	2,4	12,3	34,4	50,8	2,2	11,6	32,6	53,6	0,028
Irlandia	0,6	12,4	36,1	50,9	0,6	12,7	36,0	50,8	0,003
Niemcy	0,3	7,7	15,2	76,8	0,3	7,3	14,2	78,2	0,015
Portugalia	10,8	13,4	9,8	66,0	9,3	13,0	10,9	66,7	0,019
Szwecja	1,0	10,3	15,9	72,8	0,9	10,1	14,8	74,3	0,014
Wlk. Brytania	0,2	3,6	8,3	87,9	0,2	3,6	7,9	88,4	0,005
Włochy	14,4	23,0	20,9	41,7	11,4	23,1	21,5	44,0	0,030
Bułgaria	5,2	3,7	4,2	86,9	4,0	3,6	4,4	88,0	0,014
Cypr	30,7	23,6	16,8	29,1	30,9	22,4	16,6	30,1	0,013
Czechy	0,2	2,4	4,0	93,4	0,2	2,8	4,2	92,7	0,007
Estonia	1,8	8,3	8,6	81,3	1,7	7,9	8,4	82,0	0,007
Litwa	11,4	20,9	14,2	53,5	8,8	19,2	13,0	59,0	0,054
Łotwa	4,0	22,5	16,0	57,4	3,7	17,9	15,4	63,0	0,056
Polska	13,9	37,4	19,2	29,5	13,0	35,1	21,0	30,8	0,031
Rumunia	29,7	13,4	4,1	52,8	28,5	14,9	4,4	52,1	0,018
Słowacja	1,5	2,2	2,3	94,0	1,4	2,8	2,5	93,3	0,007
Słowenia	21,6	48,8	17,5	12,1	21,0	47,7	18,5	12,8	0,017
Węgry	6,0	9,7	10,1	74,3	5,3	9,9	10,5	74,3	0,006
Chorwacja					13,4	22,1	13,1	51,4	
Średnia zmiana* (UE-24)					0,96	1,21	1,05	2,33	

\* Średnia zmiana oznacza średnią arytmetyczną z wartości bezwzględnych różnic udziałów (w %) w roku 2010 i 2013.

Źródło: obliczenia własne na podstawie [www.europa.eu/eurostat](http://www.europa.eu/eurostat) (data utworzenia: październik 2013, grudzień 2015)

W grupie obszarowej 5-20 ha średnie zmiany wynosiły 1,21 p.p., następowały w różnych kierunkach, podobnie w grupie gospodarstw 20-50 ha, gdzie średnia

zmiana wynosiła 1,05 pp., przy czym w większości UE-15 był to spadek udziałów, natomiast w większości nowych państw członkowskich UE-12 odnotowano wzrost udziałów.

Zaobserwowane tendencje są pozytywne, szczególnie w przypadku nowych państw członkowskich, gdyż oznaczają kontynuację procesu koncentracji ziemi w gospodarstwach większych, który obserwuje się w tych krajach od czasu wejścia do UE [Babiak 2010, Poczta 2013].

Badana struktura jest silnie zróżnicowana w krajach UE (tabela 2). W 2013 roku gospodarstwa do 5 ha zajmują od 0,1% ogólnej powierzchni UR w Danii do 28,5% w Rumunii i 30,9% na Cyprze, podczas gdy odsetek ten średnio dla całej UE wynosi 6,9%. Gospodarstwa 5-20 ha zajmują od 2,8% UR na Słowacji i w Czechach do 47,7% w Słowenii (12,7% w UE), gospodarstwa 20-50 ha skupiają od 2,5% na Słowacji do 36% w Irlandii (14,2% w UE). Największe dysproporcje dotyczą gospodarstw o powierzchni co najmniej 50 ha: skupiają one od 12,8% UR w Słowenii do 93,3% na Słowacji (66% w UE).

W oparciu o metodę klasyfikacji rozmytej, przeprowadzono grupowanie krajów pod względem podobieństwa przedstawionej struktury w roku 2010 i 2013. Obliczenia zostały wykonane za pomocą autorskiego programu komputerowego, który dla danego zbioru obiektów wielowymiarowych wyznacza środki ciężkości skupisk i oblicza wartości funkcji przynależności poszczególnych obiektów do tych skupisk. Wartości początkowe stopni przynależności do klas rozmytych były ustalane losowo, co nie miało wpływu na klasyfikację końcową. Obliczenia przerywano, gdy maksimum (po klasach i po składowych) modułu różnicy wartości stopni przynależności w dwóch kolejnych iteracjach było mniejsze od 0,000001.

Zastosowana procedura grupowania doprowadziła do wyodrębnienia 4 grup krajów. Skład grup i ich charakterystykę (wartości średnie wskaźników struktury, odchylenie standardowe, współczynnik zmienności) przedstawiają tabele 3 i 4. W tabeli 3 w nawiasach podane są stopnie przynależności krajów do grup typologicznych. Kraje o najwyższym stopniu przynależności to „reprezentanci” grup, tzn. kraje najbardziej typowe dla grupy.

Grupa I obejmuje kraje o największym rozdrobnieniu badanej struktury, gdzie wysoki odsetek UR skupiony jest w gospodarstwach do 5 ha (średnio 23,2% w 2010 roku) i w gospodarstwach 5-20 ha (36,8%). Gospodarstwa 20-50 ha skupiają 18,9% UR, gospodarstwa największe - jedynie 21,2%. W 2013 roku gospodarstwa do 5 ha skupiają mniejszą część ziemi (18,8%), podobnie gospodarstwa 5-20 ha (30,6%), natomiast większa część UR jest zajmowana przez gospodarstwa największe (31,9%). W 2013 roku grupa ta powiększyła się o jeden kraj - Włochy.

Tabela 3. Przynależność krajów UE do grup typologicznych. Wartości funkcji przynależności

Grupa	2010 rok	2013 rok
I	Cypr (0,604), Grecja*(0,953), Polska (0,755), Słowenia (0,885)	Cypr (0,746), Grecja (0,590), Polska (0,915), Słowenia (0,716), Włochy (0,363)
II	Litwa (0,970), Łotwa (0,748), Portugalia (0,623), Rumunia(0,563), Włochy (0,507)	Hiszpania (0,616), Litwa (0,926), Łotwa (0,874), Portugalia (0,863), Rumunia (0,440), Chorwacja (0,580)
III	Austria (0,548), Belgia (0,851), Finlandia (0,929), Holandia (0,969), Irlandia (0,942)	Austria (0,498), Belgia (0,870), Finlandia (0,648), Holandia (0,986), Irlandia (0,911)
IV	Bułgaria (0,943), Czechy (0,885), Dania (0,975, Estonia (0,971), Francja (0,986), Hiszpania (0,462), Niemcy (0,782), Słowacja (0,867), Szwecja (0,567), Węgry (0,696), Wielka Brytania (0,970)	Bułgaria (0,957), Czechy (0,915), Dania (0,970), Estonia (0,935), Francja (0,988), Niemcy (0,709), Słowacja (0,894), Szwecja (0,448), Węgry (0,464), Wielka Brytania (0,983)

\*Dane za 2007 rok.

Źródło: obliczenia własne

Tabela 4. Wartości średnie wskaźników struktury oraz ich rozproszenie w wyodrębnionych grupach krajów UE w 2010 i 2013 roku

Grupa	Charakterystyka	Grupy obszarowe gospodarstw w ha UR							
		2010				2013			
		do 5	5-20	20-50	>=50	do 5	5-20	20-50	>=50
I	średnia	23,2	36,8	18,9	21,2	18,8	30,6	18,7	31,9
	s(x)	6,3	9,0	2,0	8,1	7,0	9,7	2,3	11,1
	V(x)	0,27	0,24	0,10	0,38	0,37	0,32	0,12	0,35
II	średnia	14,1	18,6	13,0	54,3	11,4	16,3	11,8	60,5
	s(x)	8,5	4,3	5,7	7,8	8,3	3,8	3,6	7,1
	V(x)	0,61	0,23	0,44	0,14	0,74	0,23	0,30	0,12
III	średnia	1,9	13,8	33,1	51,2	1,7	13,3	32,2	52,8
	s(x)	1,3	4,7	2,5	7,3	1,3	4,8	3,3	8,5
	V(x)	0,70	0,34	0,08	0,14	0,80	0,36	0,10	0,16
IV	średnia	2,0	6,3	9,5	82,2	1,5	5,7	8,7	84,1
	s(x)	2,1	3,1	4,4	7,6	1,7	2,8	3,9	6,6
	V(x)	1,05	0,50	0,46	0,09	1,13	0,48	0,45	0,08

Źródło: obliczenia własne

Poprawa struktury w grupie wynika częściowo ze zmian, jakie miały miejsce w krajach grupy I, a częściowo spowodowana jest przejściem Włoch do grupy I.

W II grupie gospodarstwa do 20 ha skupiają łącznie prawie o połowę mniejszą część UR, niż w grupie I i jest to odpowiednio 14,1 i 18,6% w roku 2010 i 11,4 i 16,3% w roku 2013. Gospodarstwa 20-50 ha zajmują 13% UR w roku 2010 i 11,8% w roku 2013. Gospodarstwa największe skupiają średnio 54,3% UR w 2010 roku, a w 2013 r. odsetek ten jest wyższy i wynosi 60,5%. W 2013 roku skład tej grupy zwiększył się o 2 kraje: Chorwację i Hiszpanię, natomiast Włochy w 2013 roku wchodziły w skład grupy I.

Grupa III, jako jedyna spośród wszystkich grup typologicznych nie zmieniła składu w badanych latach. Porównywalne z poprzednią grupą są udziały w powierzchni UR gospodarstw o powierzchni co najmniej 50 ha i 5-20 ha. O wyraźnej odrębności tej grupy decyduje jednak znacząco niższy odsetek UR skupiony w gospodarstwach najmniejszych, do 5 ha (nie przekracza 2%) oraz najwyższy (spośród wszystkich grup typologicznych) udział gospodarstw 20-50 ha, wynoszący około 33%.

Odmiennym rozkładem powierzchni UR cechuje się grupa IV. W krajach tej grupy ponad 80% ogółu UR skoncentrowane jest w gospodarstwach największych, o powierzchni co najmniej 50 ha. Gospodarstwa najmniejsze, do 5 ha, zajmują mniej niż 2% UR (podobnie jak w grupie III), a pozostałe gospodarstwa w sumie zajmują około 15% UR.

Największe zróżnicowanie udziałów w grupach typologicznych wykazują gospodarstwa o powierzchni co najmniej 50 ha (rozstęp wynosi powyżej 50 pp.), zaś najmniejsze – gospodarstwa do 5 ha (poniżej 20 pp.).

Wyodrębnione grupy typologiczne charakteryzują się wysokim zróżnicowaniem wewnątrzgrupowym, mierzonym wartością współczynnika zmienności  $V(x)$ , w tych kategoriach obszarowych, które w danej grupie typologicznej stanowią relatywnie niewielki odsetek, zaś niskim zróżnicowaniem w pozostałych kategoriach obszarowych.

We wszystkich grupach występowały jednakowe tendencje: spadły udziały w UR gospodarstw z grup obszarowych do 50 ha, a wzrosły gospodarstw o powierzchni co najmniej 50 ha, przy czym w grupie I i II, a więc w krajach o rozdrobnionej strukturze agrarnej były to zmiany wyraźne, natomiast w grupie III i IV – nieznaczne.

W okresie 2010-2013 zmienił się nieznacznie skład grup – tylko dwa kraje zmieniły przynależność grupową. Oznacza to, że zaobserwowany podział krajów na grupy jest utrwalony, a zmiany badanej struktury przebiegają podobnie w obrębie grup typologicznych.

Wyniki badań zaprezentowane w tabelach 3 i 4 przedstawiają podział krajów na grupy o podobnej strukturze agrarnej pod względem rozkładu powierzchni UR według grup obszarowych. Inny podział otrzymuje się w grupowaniu pod względem rozkładu liczby gospodarstw (tabela 5), co przedstawiono w pracy [Bożek 2016]. Jednakże w większości przypadków kraje, w których jest

największy odsetek gospodarstw małych obszarowo, należą do grup, gdzie największy odsetek ziemi skupiają gospodarstwa małe.

Tabela 5. Grupy krajów UE o podobnym rozkładzie liczby gospodarstw według grup obszarowych UR w 2013 roku

Grupa	Skład grupy	Grupy obszarowe gospodarstw w ha UR			
		do 5	5-20	20-50	>=50
I	Bułgaria, Cypr, Grecja, Portugalia, Rumunia, Węgry.	83,7	11,3	2,8	2,2
II	Hiszpania, Litwa, Łotwa, Polska, Słowacja, Słowenia, Włochy, Chorwacja.	56,2	30,5	7,5	5,8
III	Austria, Estonia, Holandia	30,3	35,3	20,8	13,6
IV	Belgia, Czechy, Dania, Finlandia, Francja, Irlandia, Niemcy, Szwecja, Wielka Brytania.	11,8	33,7	24,8	30,1

Źródło: opracowanie własne na podstawie [Bożek 2016]

Tak więc kraje należące do grup o rozdrobnionej strukturze pod względem rozkładu liczby gospodarstw (grupa I i II), należą również do grup o rozdrobnionej strukturze pod względem rozkładu UR (I i II). Wyjątek stanowią Bułgaria i Słowacja. W tych krajach odsetek gospodarstw do 5 ha wynosi kolejno: 86,9%, 58,9%, a skupiają one odpowiednio: 4%, 1,4% użytków rolnych. Jednocześnie gospodarstwa największe stanowią 3,6% w Bułgarii i 13,1% na Słowacji, zajmując odpowiednio 88% i 93,3% UR, stąd kraje te w grupowaniu pod względem rozkładu powierzchni UR należą do grupy IV, o najlepszej strukturze, natomiast pod względem rozkładu liczby gospodarstw w grupach obszarowych należą do grup krajów o największym rozdrobnieniu.

## PODSUMOWANIE

Przemiany struktury obszarowej gospodarstw rolnych w latach 2010-2013 są kontynuacją tendencji z lat 2002-2010: zmniejsza się udział gospodarstw poniżej 20 ha w użytkowanej ziemi, a zwiększa się udział gospodarstw dużych, o powierzchni co najmniej 50 ha. W krajach o rozdrobnionej strukturze agrarnej (większość UE-12) tendencja wzrostowa powierzchni UR występuje także w gospodarstwach 20-50 ha, natomiast w krajach UE-15 w tej grupie obszarowej następuje spadek powierzchni UR.

Nadal występuje bardzo duże zróżnicowanie struktury obszarowej gospodarstw rolnych w krajach UE. Można wyodrębnić 4 grupy krajów podobnych pod względem rozkładu powierzchni użytków rolnych według grup wielkościowych gospodarstw. Kraje należące do grupy I (Cypr, Grecja, Polska, Słowenia, Włochy) charakteryzuje największe rozdrobnienie badanej struktury: gospodarstwa do 5 ha



zajmują w 2013 roku średnio 18,8% powierzchni UR, 30,6% UR przypada na gospodarstwa 5-20 ha, 18,7% UR skupiają gospodarstwa 20-50 ha, a gospodarstwa o powierzchni co najmniej 50 ha zajmują średnio tylko 31,9% UR. W krajach grupy II (Hiszpania, Litwa, Łotwa, Portugalia, Rumunia, Chorwacja), gdzie rozdrobnienie jest mniejsze, gospodarstwa z dwóch pierwszych grup obszarowych skupiają znacząco mniejszy odsetek UR: odpowiednio 11,4% i 16,3%, a na gospodarstwa duże i bardzo duże przypada 11,8 i 60,5% UR. Inaczej rozkłada się powierzchnia UR w krajach grupy III (Austria, Belgia, Finlandia, Holandia, Irlandia), gdzie gospodarstwa do 20 ha skupiają niewielką część powierzchni UR: odpowiednio 1,7% i 13,3%, a kolejne dwie grupy obszarowe zajmują odpowiednio 32,2% i 52,8%. W krajach grupy IV (Bułgaria, Czechy, Dania, Estonia, Francja, Niemcy, Słowacja, Szwecja, Węgry, Wielka Brytania) średnio 84,1% UR skupiają gospodarstwa o powierzchni co najmniej 50 ha, a trzy pierwsze grupy obszarowe zajmują odpowiednio: 1,5%, 5,7% i 8,7% UR.

Zaobserwowany podział krajów na grupy jest utrwalony (w okresie 2010-2013 tylko dwa kraje zmieniły przynależność grupową). Struktura grup I i II w badanym okresie poprawiła się: spadły udziały gospodarstw z grup obszarowych do 50 ha w zajmowanej powierzchni UR, a wzrosły gospodarstw o powierzchni co najmniej 50 ha.

## BIBLIOGRAFIA

- Babiak J. (2010) Zmiany w strukturze rolnictwa krajów Unii Europejskiej. *Roczniki Integracji Europejskiej*, 4, 87-97.
- Bożek J. Bożek B. (2011) Typologia struktury agrarnej województw w ujęciu dynamicznym z zastosowaniem klasyfikacji rozmytej. *Metody ilościowe w badaniach ekonomicznych*, XII/2, 91-100.
- Bożek J. (2013) Klasyfikacja podregionów pod względem podobieństwa struktury agrarnej. *Wiadomości Statystyczne*, 9, 1-16.
- Bożek J. (2016) Transformation of Agrarian Structure of EU Countries within the Period of 2010–2013. *Acta Scientiarum Polonorum, Oeconomia*, 15 (2), 15-25.
- Farm Structure Survey 2013 – Main Results (2015): Main Tables and Database. Eurostat.
- Jajuga K. (1984) Zbiory rozmyte w zagadnieniu klasyfikacji. *Przegląd Statystyczny*, 31(3-4), 237-250.
- Klepacki B., Żak A. (2013) Przemiany agrarne na terenach polskich przed i po integracji z Unią Europejską. *Journal of Agribusiness and Rural Development*, 4(30), 1-17.
- Kukuła K. (1989) Statystyczna analiza strukturalna i jej zastosowanie w sferze usług produkcyjnych dla rolnictwa. *Zeszyty Naukowe AE w Krakowie, Seria specjalna: Monografie*, 89, Kraków.
- Poczta W. (red.) (2013) *Gospodarstwa rolne w Polsce na tle gospodarstw Unii Europejskiej - wpływ WPR*. GUS, Warszawa.
- Stańko S., Mikuła A. (2016) Zmiany struktury obszarowej gospodarstw rolnych w krajach UE-15 i w Polsce. *Zeszyty Naukowe SGGW w Warszawie, Problemy Rolnictwa Światowego*, XXXI/16 (1), 234-244.

Zadeh Lofti A. (1965) Fuzzy Sets. Information and Control, 8, 338-353.

**DIVERSIFICATION AND DYNAMICS OF CHANGES  
IN THE UTILISED AGRICULTURAL AREA BY FARMS AREAL  
GROUPS IN THE EUROPEAN UNION COUNTRIES  
IN THE YEARS 2010-2013**

**Abstract:** The present paper discusses the diversification and dynamics of changes in the utilised agricultural area in farms representing different areal groups in the EU in the years 2010-2013. The study has been conducted based on the Eurostat data. The following areal groups have been taken into account: up to 5 ha, 5-20, 20-50 and over 50 ha of UAA. Using the selected methods of statistical analysis of structures, the authors determined the scale and direction of changes of the analysed structure in particular countries. Based on the fuzzy classification, four groups of countries with similar percentage of UAA by different areal groups have been identified.

**Keywords:** utilised agricultural area, European Union, dynamics, fuzzy classification

## AKTYWNOŚĆ ZAWODOWA KOBIET NA POLSKIM RYNKU NIERUCHOMOŚCI NA PRZYKŁADZIE ZAWODÓW REGULOWANYCH

**Iwona Forys (ORCID: 0000-0002-2294-0672)**

Wydział Nauk Ekonomicznych i Zarządzania

Uniwersytet Szczeciński

e-mail: forys@wneiz.pl

**Streszczenie:** Ostatnia dekada przyniosła zwiększenie współczynnika feminizacji na polskim rynku pracy. Celem badania jest ocena stopnia oraz zakresu zmian w czasie feminizacji licencjonowanych zawodów na polskim rynku nieruchomości. Analizowano współczynnik feminizacji zawodów regulowanych rynku nieruchomości w czasie i w ujęciu przestrzennym. Najmniej sfeminizowanym zawodem jest rzeczoznawca majątkowy, najsilniej pośrednik w obrocie nieruchomościami. Wskazano, że największe wahania zmian struktur kobiet z uprawnieniami w tych zawodach miało miejsce w 2007 oraz 2013 roku. Wyniki wskazują na niemonotoniczne zmiany w strukturze kobiet uzyskujących licencję w latach 1999–2013 obowiązywania egzaminów państwowych.

**Słowa kluczowe:** rynek nieruchomości, współczynnik feminizacji, zawody regulowane

### WSTĘP

Na rynku pracy można zauważyć zjawisko feminizacji, które z różną siłą i natężeniem dotyczą poszczególnych sektorów gospodarki [Horodeński, Sadowska–Snarska 2003]. Na stan ten wpływają zarówno czynniki gospodarcze, społeczne czy kulturowe danego kraju. Zjawisko zwiększonej reprezentacji kobiet w pewnych branżach jest określane jako horyzontalna segregacja zawodowa [Zwiech 2010]. Najbardziej sfeminizowane branże w UE to edukacja, ochrona zdrowia i opieka społeczna. W Polsce w 2010 roku najbardziej sfeminizowanymi sektorami gospodarki narodowej były: ochrona zdrowia i opieka społeczna, gdzie na 550 kobiet przypadało 110 mężczyzn oraz edukacja - 575 kobiet na 153

mężczyzn [Kobieta i mężczyzna ... 2012]. Wysoki udział kobiet notowany jest również w pośrednictwie finansowym, hotelarstwie oraz gastronomii. Oprócz wspomnianych zawodów, w Polsce w ostatnich dwóch dekadach można również zauważyć zwiększoną aktywność kobiet na rynku nieruchomości po stronie profesjonalnej obsługi rynku, w szczególności w wycenie nieruchomości, zarządzaniu nieruchomościami oraz pośrednictwie w obrocie nieruchomościami, co potwierdza liczba uzyskiwanych licencji i uprawnień zawodowych.

Do końca 2013 roku zawody rzeczoznawcy majątkowego, pośrednika w obrocie nieruchomościami oraz zarządcy nieruchomości były w Polsce regulowane. Ich wykonywanie wymagało uzyskania stosownych uprawnień zawodowych nadawanych przez ministra właściwego do spraw budownictwa, po pozytywnym zdaniu egzaminów państwowych. Osoby, które uzyskały uprawnienia rzeczoznawcy majątkowego lub licencję pośrednika w obrocie nieruchomościami lub licencję zarządcy nieruchomości były wpisywane do Centralnego Rejestru tych zawodów, prowadzonego przez Ministra. Wraz ze zmianą przepisów Ustawy o gospodarce nieruchomościami zmieniały się wymagania w stosunku do poszczególnych zawodów, które następowały w kilku fazach:

- rzeczoznawca majątkowy: lata 1991–1997, 1998–2013 oraz ostatecznie w 2014 roku – zmniejszenie wymagań szkoleniowych i egzaminacyjnych,
- zarządca nieruchomości: lata 1998–2006, 2007–2008, 2009–2013 (brak egzaminu), oraz 1 styczeń 2014 r. deregulacja zawodu,
- pośrednik w obrocie nieruchomościami: lata 1998–2006, 2007–2008, 2009–2013 (brak egzaminu) oraz 1 styczeń 2014 roku deregulacja zawodu.

Celem badania jest ocena stopnia oraz zakresu zmian w czasie feminizacji licencjonowanych zawodów na polskim rynku nieruchomości, również w ujęciu przestrzennym. W badaniu weryfikowano hipotezę o rosnącym udziale kobiet w trzech licencjonowanych (do końca 2013 roku) zawodach na polskim rynku nieruchomości. W tym celu wykorzystano dane z Centralnego Rejestru prowadzonego przez Ministra Budownictwa, dotyczące liczby uzyskiwanych uprawnień i licencji zawodowych nadawanych przez właściwego ministra w latach 1990–2013, w podziale na płeć. Analizowano współczynnik feminizacji poszczególnych zawodów oraz wyznaczono wskaźniki zróżnicowania struktury uzyskiwanych uprawnień i licencji zawodowych w Polsce w latach 1998–2013.

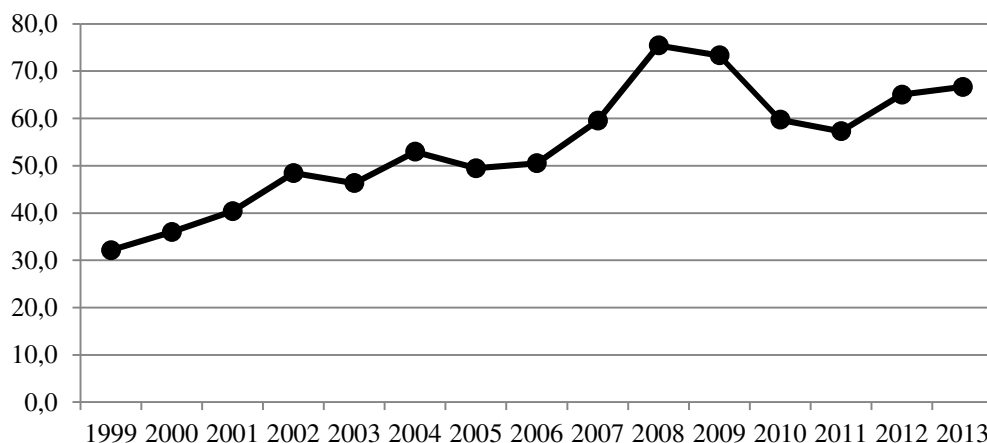
### **Współczynnik feminizacji zawodów regulowanych rynku nieruchomości**

Współczynnik feminizacji stanowi odsetek kobiet, którym nadano uprawnienia rzeczoznawcy majątkowego oraz licencję pośrednika w obrocie nieruchomościami lub zarządcy nieruchomości. Należy przy tym wspomnieć, że przystąpienie do egzaminu państwowego w latach 1999–2013 wymagało od kandydata ukończenia studiów podyplomowych w jednej z tych trzech specjalności oraz odbycia praktyki zawodowej. Program studiów podyplomowych w znacznej części obejmował wiedzę prawniczą, ekonomiczną oraz techniczną. W niewielkim

zakresie dotyczył kompetencji miękkich, wiedzy socjologicznej czy psychologii. Interdyscyplinarny charakter kompetencji zawodów rynku nieruchomości, ogromny zakres materiału do przyswojenia powodował, że większość słuchaczy studiów podyplomowych legitymowała się wykształceniem wyższym w jednej z tych trzech podstawowych dziedzin wiedzy, uzupełniając ją o niezbędną wiedzę z pozostałych dwóch obszarów oraz szczegółowe kompetencje właściwe czynnościom wykonywanym na rynku nieruchomości.

Najdłużej na polskim rynku funkcjonują uprawnienia rzeczoznawcy majątkowego (od połowy 1991 roku). Jednak z uwagi na porównywalność omawianych trzech zawodów do analizy przyjęto lata 1999–2013. W tym okresie uprawnienia zawodowe uzyskało 5 801 osób, w tym 42% stanowiły kobiety (2 436 uprawnień). Jednak odsetek kobiet uzyskujących uprawnienia zawodowe do szacowania wartości nieruchomości nie był na jednakowym poziomie (rysunek 1).

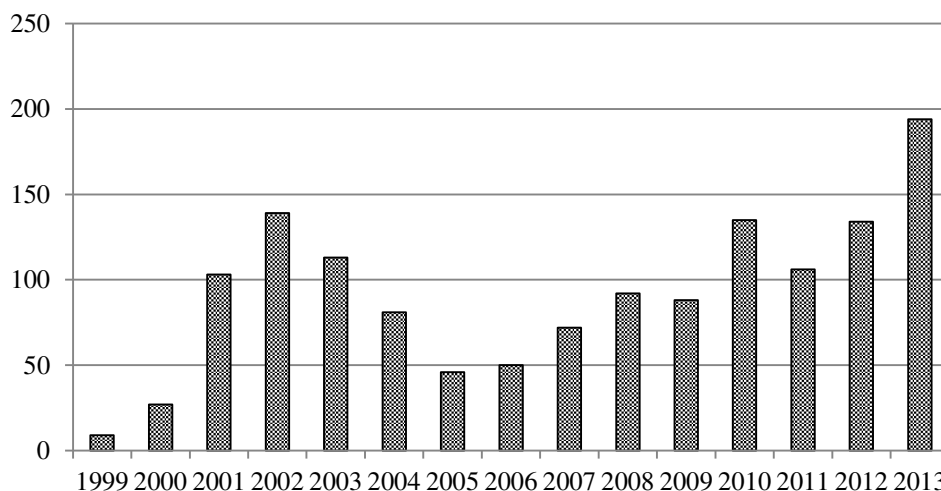
Rysunek 1. Dynamika współczynnika feminizacji uprawnień rzeczoznawcy majątkowego wydanych w latach 1999 – 2013 (%)



Źródło: opracowanie własne

W badanych latach współczynnik feminizacji zawodu rzeczoznawcy majątkowego charakteryzował się: tendencją rosnącą w latach 1999–2008, spadkową w latach 2009–2011 oraz ponownym wzrostem w latach 2012–2013. Od 2006 roku ponad połowa uzyskanych uprawnień należała do kobiet. Najwięcej uprawnień zawodowych kobiety uzyskały w 2013 roku (rysunek 2).

Rysunek 2. Dynamika liczby uprawnień rzeczoznawcy majątkowego wydanych kobietom w latach 1999 – 2013

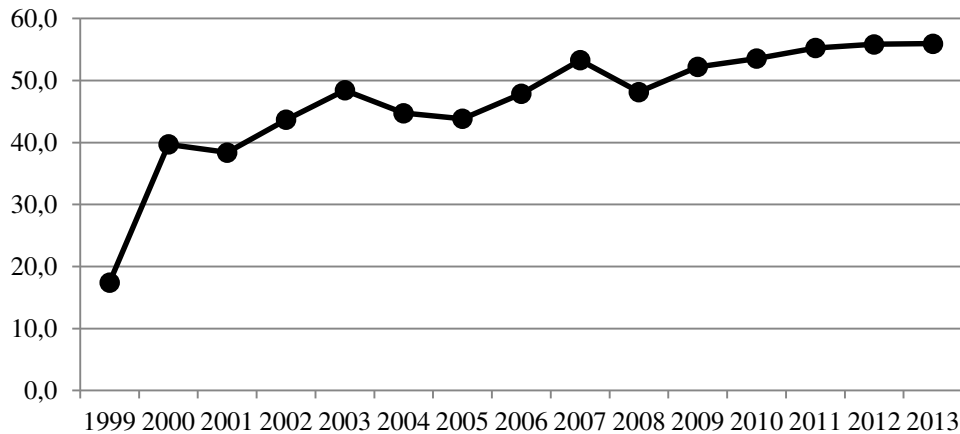


Źródło: opracowanie własne

Najmniej uprawnień uzyskiwały kobiety w latach 1999–2000, kiedy zawód rzeczoznawcy majątkowego dopiero rozwijał się na polskim rynku, a w ustawie znajdował się wymóg wykształcenia technicznego, prawniczego, rolniczego lub ekonomicznego. W kolejnych okresach wahania udziału kobiet w liczbie osób uzyskujących uprawnienia zawodowe rzeczoznawcy majątkowego były również skorelowane z koniunkturą na rynku nieruchomości, która oznaczała dużą liczbę transakcji rynkowych, finansowanie kredytem zakupu nieruchomości, co generowało popyt na usługi rzeczoznawców majątkowych.

W przypadku zarządcy nieruchomości udział kobiet uzyskujących uprawnienia w ostatnich latach badania (2009–2013) jest stabilny przekracza 50%, natomiast w latach 2001–2007 przekraczał 40% (rysunek 3).

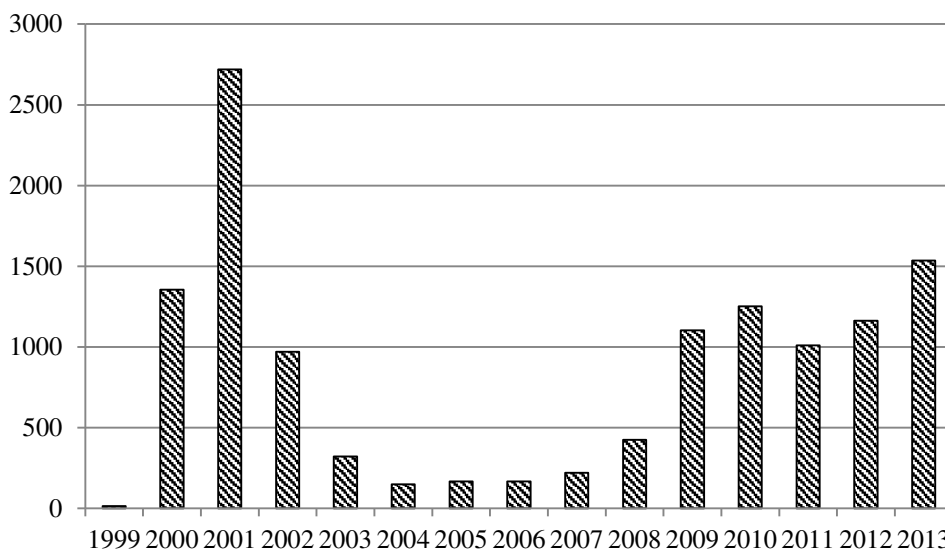
Rysunek 3. Dynamika współczynnika feminizacji uprawnień rzeczoznawcy majątkowego wydanych w latach 1999 – 2013 (%)



Źródło: opracowanie własne

W całym analizowanym okresie współczynnik feminizacji zawodu zarządcy nieruchomości wyznaczany na podstawie liczby uzyskanych licencji zawodowych wyniósł 46,7%. Najwięcej licencji zarządcy uzyskiwały kobiety w tzw. trybie uproszczonym na podstawie dotychczasowego dorobku zawodowego, bez zdawania egzaminów w latach 2000–2001 (rysunek 4). Najmniej w latach 2004–2007, kiedy kryteria egzaminacyjne były zaostrzone a praktyka zawodowa wymagała sporządzenia trzech planów zarządzania nieruchomością [Bryx 2007]. Należy jednak zwrócić uwagę, że jest to zawód najmocniej osadzony w kompetencjach technicznych, budownictwie, wiedzy z zakresu infrastruktury technicznej budynku, systemów inteligentnych budynku. Dodatkowo należy zwrócić uwagę na duży zasób nieruchomości mieszkalnych tzw. wspólnot mieszkaniowych, którymi zarządzają kobiety w ramach małych przedsiębiorstw lub jako osoby zatrudnione na etatach [Foryś, Nowak 2012].

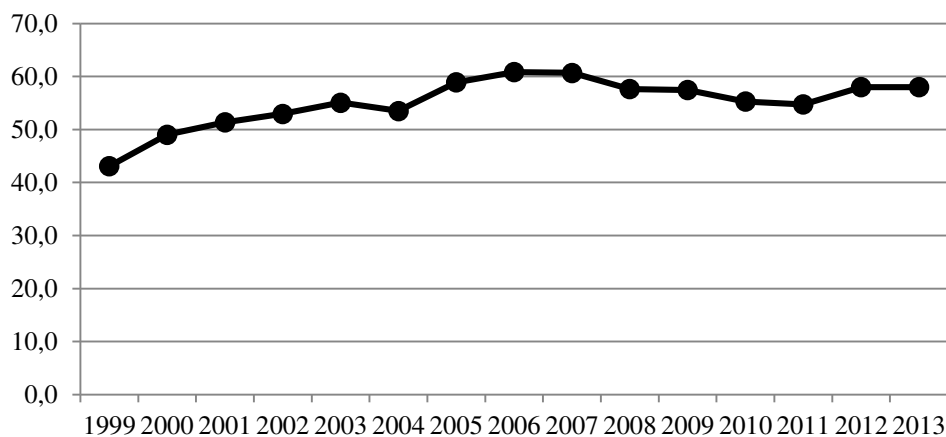
Rysunek 4. Dynamika liczby licencji zarządcy nieruchomości wydanych kobietom w latach 1999 – 2013



Źródło: opracowanie własne

Najbardziej sfeminizowanym zawodem spośród trzech omawianych jest zawód pośrednika w obrocie nieruchomościami. Współczynnik począwszy od 2001 roku przekracza w badanym okresie 50%. W przypadku tej licencji widoczne są dwa istotne wzrosty współczynnika feminizacji, które pokrywają się z koniunkturą na rynku nieruchomości: lata 1999–2001 oraz lata 2004–2009 (rysunek 5).

Rysunek 5. Dynamika współczynnika feminizacji licencji pośrednika w obrocie nieruchomościami wydanych w latach 1999 – 2013 (%)

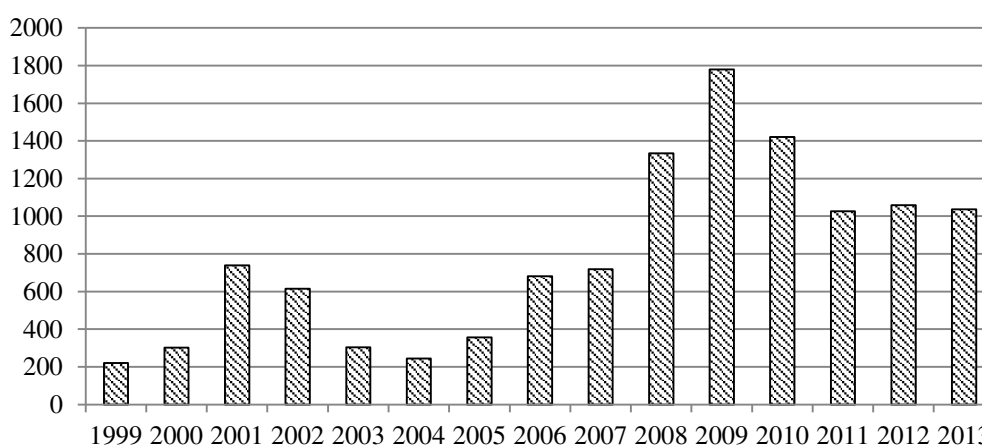


Źródło: opracowanie własne



Wykonywanie zawodu pośrednika w obrocie nieruchomościami wymaga licznych kompetencji miękkich, zrozumienia potrzeb klienta, jego preferencji i gustu w doborze odpowiedniej oferty na rynku nieruchomości [Foryś 2009, Kalinowska–Nawrotek 2005]. Być może jest to przyczyna większego niż mężczyzn zainteresowania kobiet tym zawodem. Najwięcej licencji kobiety uzyskały w okresie zmniejszenia wymagań egzaminacyjnych, ale również pokrywającym się z dekoniunktura gospodarczą [Kurzynowski 2001], co może oznaczać przekwalifikowanie się kobiet w związku z utratą dotychczasowej pracy (rysunek 6).

Rysunek 6. Dynamika liczby licencji pośrednika w obrocie nieruchomościami wydanych kobietom w latach 1999 – 2013



Źródło: opracowanie własne

We wszystkich trzech zawodach widoczna jest wzmożona aktywność kobiet w uzyskiwaniu uprawnień i licencji w latach 2008–2013. W tym okresie w każdym z zawodów ponad połowę stanowiły kobiety, a tendencja współczynnika feminizacji tych zawodów jest stale rosnąca mimo, iż z mniejszą dynamiką.

### Badanie zróżnicowania struktur

Do oceny zmian struktury kobiet z uprawnieniami do wskazanych zawodów w czasie wykorzystano proste miary zróżnicowania struktur oparte na relacjach udziałów danej składowej w strukturze w okresie  $t$  oraz  $t+1$ . Wartości miar zróżnicowania struktur znajdują się w przedziale  $<0;1>$ , przy czym słabsze zróżnicowanie struktury oznacza wartość wskaźnika bliską zera, natomiast silne zróżnicowanie struktury wartość wskaźnika bliską jedności. Proponowane w literaturze miary różnią się reakcją na zmiany składowych o różnych udziałach [Kukuła 1996, Walesiak 1983]. W pierwszym kroku wyznaczono trzy miary, które w różnym stopniu reagują na udziały składowych struktury (wzory 1-3).

Miara równomiernie reagująca na zmiany w poziomie składowych o różnych udziałach:

$$V_1 = \frac{\sum_{i=1}^k |\alpha_{it} - \alpha_{i(t+1)}|}{2}. \quad (1)$$

Współczynnik dywergencji Clarka to miara, która silniej reaguje na odchylenia składowych o niskim udziale:

$$V_2 = \left[ \frac{1}{k} \sum_{i=1}^k \left( \frac{\alpha_{it} - \alpha_{i(t+1)}}{\alpha_{it} + \alpha_{i(t+1)}} \right) \right]. \quad (2)$$

Współczynnik zaproponowany przez M. Walesiaka to miara, która silniej reaguje na odchylenia składowych o wysokim udziale:

$$V_3 = \left[ \frac{1}{2} \sum_{i=1}^k |\alpha_{it}^2 - \alpha_{i(t+1)}^2| \right]^{\frac{1}{2}}, \quad (3)$$

gdzie w powyższych wzorach:

$\alpha_{it}$  – udział  $i$ -tej składowej w okresie poprzedzającym badany,

$\alpha_{i(t+1)}$  – udział  $i$ -tej składowej w badanym okresie,

$i = 1, 2, \dots, k$  – kolejne składowe struktury,

$t = 0, 1, \dots, m$  – kolejne lata w badanym okresie.

Natomiast wykorzystując jako podstawę badania okres bazowy, można sprawdzić monotoniczność zmian zachodzących w danej strukturze wykorzystując miarę monotoniczności struktury postaci [Kukuła 1996]:

$$\eta_m = \frac{\sum_{i=1}^k |\alpha_{im} - \alpha_{i0}|}{\sum_{t=0}^{m-1} \sum_{i=1}^k |\alpha_{it} - \alpha_{i(t+1)}|}. \quad (4)$$

W celu określenia zróżnicowania struktur badano strukturę uzyskiwanych licencji pośrednika w obrocie nieruchomościami i zarządcy nieruchomości oraz uprawnień zawodowych rzeczoznawcy majątkowego przez kobiety i mężczyzn. Uzyskane wyniki wskazują na niemonotoniczne zmiany w strukturze, a wartości bliskie zeru wskazują na bardzo słabe zróżnicowanie struktur (tabela 1).

Tabela 1. Miary zróżnicowania struktur zawodów regulowanych rynku nieruchomości w Polsce

Lata	Rzeczoznawca majątkowy			Pośrednik w obrocie nieruchomościami			Zarządca nieruchomości		
	Miary zróżnicowania			Miary zróżnicowania			Miary zróżnicowania		
	V1	V2	V3	V1	V2	V3	V1	V2	V3
2000	0,039	0,037	0,196	0,042	0,035	0,210	0,118	0,091	0,401
2001	0,044	0,039	0,210	0,020	0,017	0,142	0,033	0,038	0,172
2002	0,080	0,067	0,283	0,019	0,015	0,128	0,050	0,039	0,230
2003	0,021	0,017	0,145	0,019	0,030	0,135	0,042	0,034	0,206
2004	0,066	0,054	0,257	0,035	0,040	0,194	0,023	0,021	0,154
2005	0,034	0,028	0,184	0,037	0,04	0,189	0,025	0,026	0,154
2006	0,010	0,008	0,100	0,011	0,801	0,096	0,048	0,038	0,223
2007	0,090	0,800	0,500	0,017	0,022	<b>0,511</b>	0,053	<b>0,802</b>	<b>0,527</b>
2008	0,159	0,157	0,000	0,017	0,011	0,000	0,047	0,039	0,000
2009	0,021	0,025	0,000	0,012	0,015	0,000	0,027	0,025	0,000
2010	0,136	0,131	0,000	0,014	0,017	0,000	0,012	0,012	0,000
2011	0,024	0,021	0,000	0,200	0,022	0,000	0,012	0,011	0,000
2012	0,077	0,068	0,000	0,017	0,022	0,000	0,004	0,004	0,000
2013	0,017	0,016	0,000	0,290	<b>0,577</b>	0,000	0,280	<b>0,577</b>	0,000

Źródło: obliczenia własne

Wyjątek stanowi 2007 oraz 2013 rok w przypadku zarządcy nieruchomości oraz pośrednika w obrocie nieruchomościami. W 2007 roku miara V3, która silniej reaguje na odchylenia składowych o wysokim udziale ma wartość znacznie wyższą niż w pozostałych latach (0,511) dla obu zawodów, na co mogły mieć wpływ ułatwienia egzaminacyjne począwszy od 2009 roku, które zachwiały dotychczasową tendencję. W przypadku zarządców zauważono bardzo wysoką wartość miary V2, która silniej reaguje na odchylenia składowych o niskim udziale (0,802). Podobnie w 2013 roku można zauważyć wyższe wartości miary V3 w przypadku pośredników w obrocie nieruchomościami (0,577) oraz miary V2 w przypadku zarządców nieruchomości (0,577). W obu przypadkach był to ostatni rok obowiązywania egzaminów państwowych.

Dodatkowo wyznaczono miarę monotoniczności struktury przyjmując jako podstawę badania okres bazowy – 1999 rok (według wzoru 4). W tabeli 2 zestawiono wyniki obliczeń dla poszczególnych zawodów rynku nieruchomości w Polsce w latach obowiązywania licencji i uprawnień zawodowych.

Tabela 2. Miara monotoniczności struktury zawodów regulowanych rynku nieruchomości w Polsce (rok bazowy 1999)

Zawód	Miara monotoniczności	Wniosek
Rzeczoznawcy	0,00008	niemotoniczne zmiany w strukturze
Pośrednicy	0,00000	niemotoniczne zmiany w strukturze
Zarządcy	0,000002	niemotoniczne zmiany w strukturze

Źródło: obliczenia własne

Uzyskane wyniki wskazują na niemonotoniczne zmiany w strukturze kobiet uzyskujących licencję pośrednika w obrocie nieruchomościami czy zarządcy nieruchomości lub uprawnienia zawodowe rzeczoznawców majątkowych w latach 1999–2013 obowiązywania egzaminów państwowych.

## WNIOSKI

Dalsze badania są możliwe jedynie w przypadku zwozu rzeczoznawcy majątkowego, który nie został zderegulowany od 1 stycznia 2014 roku i nadal prowadzony jest Centralny Rejestr.

Przeciętny współczynnik feminizacji dla poszczególnych zawodów w badanych latach 1999–2013 był zróżnicowany, podobnie jak i tendencja oraz jej dynamika udziału kobiet w nabywaniu uprawnień zawodowych na poszczególnych rynkach wojewódzkich:

Rzeczoznawca majątkowy:	42,0%,
Zarządca nieruchomości:	46,7%,
Pośrednik w obrocie nieruchomościami:	56,1%.

Przy czym najbardziej sfeminizowanym zawodem jest grupa pośredników w obrocie nieruchomościami. Wynika to również z charakteru pracy, nienormowanego czasu pracy i predyspozycji kobiet w nawiązywaniu i utrzymywaniu kontaktów z nabywcami mieszkań i domów, które stanowią ponad 60% transakcji na rynku [Buć 2009]. Natomiast współczynnik zróżnicowania w każdym badanym przypadku nie wskazywał na monotoniczność struktury, co sugerowałoby potencjalne zmiany w strukturze analizowanego zjawiska.

Dalszy kierunek badań będzie miał na celu zweryfikowanie czy tendencja zaobserwowana w Polsce ma również miejsce w innych krajach unii Europejskiej oraz w wiodących gospodarkach światowych. Kontynuacja badań na rynku polskim będzie utrudniona faktem deregulacji omawianych zawodów a tym samym braku dalszych wpisów do rejestru osób uzyskujących kompetencje w zawodzie pośrednika w obrocie nieruchomościami oraz zarządcy nieruchomości.

## BIBLIOGRAFIA

- Bryx M. red. (2007) Wprowadzenie do zarządzania nieruchomościami. Wydawnictwo Poltext, Warszawa.
- Buć M. (2009) Wpływ zaprogramowania kulturowego na stopień maskulinizacji społeczeństwa polskiego. [w]: Winclawska M., Brodzińska B. Płeć w życiu publicznym. Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, Toruń, 55-66.
- Foryś I. red. (2009) Obrót nieruchomościami. Wydawnictwo Poltext, Warszawa.
- Foryś I., Nowak M. (2012) Spółdzielnia czy wspólnota? Zarządzanie zasobami mieszkaniowymi. Wydawnictwo Poltext, Warszawa.

- Horodeński R., Sadowska-Snarska C. (2003) Rynek pracy w Polsce na progu XXI wieku. Aspekty makroekonomiczne i regionalne. Wyższa Szkoła Ekonomiczna, Białystok-Warszawa.
- Kalinowska - Nawrotek B. (2005) Dyskryminacja kobiet na polskim rynku pracy. Akademia Ekonomiczna, Poznań.
- Kobieta i mężczyzna na rynku pracy (2012) Wydawnictwo Głównego Urzędu Statystycznego w Warszawie.
- Kukuła K. (1996) Statystyczne metody analizy struktur ekonomicznych. Wydawnictwo Edukacyjne, Kraków.
- Kurzynowski A. (2001) Sytuacja społeczno – zawodowa bezrobotnych kobiet. Bariery i stimulatory ich aktywizacji zawodowej. Szkoła Główna Handlowa, Warszawa.
- Ustawa z dnia 21 sierpnia 1997 r. o gospodarce nieruchomościami (Dz. U. z 2015 r. poz. 782 z zm.).
- Walesiak M. (1983) Propozycje miar odległości struktur udziałowych. Wiadomości Statystyczne, 10.
- Zwiech P. (2010) Segregacja zawodowa kobiet w Polsce. [w]: Gawrycka M., Wasilczuk J., Zwiech P. Szklany sufit i ruchome schody – kobieta na rynku pracy. CeDeWu, Warszawa, 29-44.

## SUMMARY

### PROFESSIONAL ACTIVITY OF POLISH WOMEN ON THE REAL ESTATE MARKET. THE EXAMPLE OF REGULATED PROFESSIONS

**Abstract:** Today's labour market sees a trend of feminisation which is developing with varying intensity in individual segments of economy. Its shape in individual countries is influenced by local economic, social and cultural environment. The phenomenon strengthened representation of women in certain industries is referred to as horizontal occupational segregation (Zwiech 2010). In the EU the most feminized industries are education, health and social care. The aim of this study is to assess the level and range of changes over time in the process of feminisation of regulated professions on the Polish real estate market. With this aim in view, I used data on the number of professional licenses granted by the relevant Minister between 1990–2014 broken down by gender and geographical location (voivodships). The study verifies a hypothesis of growing women's participation in the Polish real estate market in three professions that had been regulated before 2014.

**Keywords:** real estate market, feminization coefficient, regulated professions

## WŁASNOŚCI FUNKCJI WARTOŚCI DLA STOCHASTYCZNEGO PROBLEMU STEROWANIA OPTYMALNEGO TYPU MAYERA

**Wiesław Grygierzec**

Wydział Rolniczo-Ekonomiczny  
Uniwersytet Rolniczy w Krakowie  
e-mail: rrgrygie@cyf-kr.edu.pl

**Streszczenie:** Niniejszy artykuł jest kontynuacją rozważań dotyczących problemu stochastycznego sterowania optymalnego w tzw. przypadku Mayera. Problemy takie opisywane są poprzez stochastyczne równanie różniczkowe typu Ito a funkcjonal kosztu jest zależny od stanu układu w czasie końcowym. Jest to w szczególności model dyfuzyjny, modele takie są adekwatne do opisu zjawisk biologicznych i ekonomicznych w których z przyczyn naturalnych mamy do czynienia z oddziaływaniem dużej ilości niezależnych sił losowych. Problem sterowania optymalnego polega na podejmowaniu na podstawie możliwie najnowszych informacji, odpowiednich decyzji spośród wszystkich możliwych w celu osiągnięcia zamierzonego celu co realizuje się poprzez minimalizację funkcjonału kosztu. Ważną rolę odgrywa tutaj tzw. funkcja wartości. W niniejszym artykule autor udawania kolejne własności funkcji wartości dla tzw. problemu Mayera czyli dla specjalnej postaci funkcjonału kosztu.

**Słowa kluczowe:** stochastyczne sterowanie optymalne, funkcja wartości, problem Mayera

### WSTĘP

Rozważamy problem stochastycznego sterowania optymalnego w czasie ciągłym. W najogólniejszym ujęciu składają się na niego:

- Po pierwsze: *układ dynamiczny*, którego stan  $x(t, \omega)$  zmienia się w czasie, ewoluujący w losowym środowisku, formalnie ujętym jako przestrzeń probabilistyczna  $(\Omega, \mathcal{F}, P)$ , czyli ogół wszystkich możliwych scenariuszy czynników losowych. Stan układu jest reprezentowany przez wektor liczb

potrzebny do opisu problemu i oznaczany jako - stan w chwili  $t$  przy scenariuszu  $\omega \in \Omega$ . Definiujemy dynamikę stanu układu, tzn. odwzorowanie:

$$[0, T] \ni t \rightarrow x(t, \omega),$$

dla każdego  $\omega \in \Omega$ , poprzez proces stochastyczny albo stochastyczne równanie różniczkowe.

- Po drugie: *sterowanie* albo inaczej *strategia*. Dynamika  $t \rightarrow x(t)$  jest dodatkowo zależna od parametru którym jest sterowanie  $u = u(t, \omega)$  a więc pewien proces którego wartość jest obliczana dla każdego  $t$  na podstawie informacji dostępnych do chwili  $t$ , bez posiadanej wiedzy o przyszłości. Sterowania powinny spełniać pewne ograniczenia i zbiór takich sterowań oznaczamy  $\mathcal{U}$  - zbiór sterowań dopuszczalnych.
- Po trzecie: *kryterium kosztu* lub *zysku*. Zadanie polega na minimalizacji (lub maksymalizacji) funkcjonału  $J(x, u)$  po zbiorze wszystkich dopuszczalnych sterowaniach

$$J(x, u) \rightarrow \min_{u \in \mathcal{U}},$$

gdzie najogólniej rozważamy funkcjonał

$$J(x, u) = E \left[ \int_0^T l(x(t, \omega; u(t))) dt + g(x(T, \omega; u(T))) \right], \quad (1)$$

gdzie  $l$  jest tzw. kosztem bieżącym natomiast  $g$  jest kosztem końcowym. Czas  $T$  może być albo deterministyczny albo losowy tzn.  $T = \tau(\omega)$  czyli tzw. *moment stopu*. Szczególną rolę odgrywa tzw. *funkcja wartości* czyli minimum funkcjonału kosztu

$$v = \min_{u, \tau} J(x, u, \tau). \quad (2)$$

Funkcja ta posiada pewne własności które pozwalają w pewnych sytuacjach na znalezienie rozwiązania problem optymalnego sterowania stochastycznego czyli na znalezieniu sterowania  $u$  i (lub) momentu stopu  $\tau$ .

## PROBLEM STEROWANIA OPTYMALNEGO MAYERA

**Stochastyczny układ ze sterowaniem**

Rozważamy następujący stochastyczny układ ze sterowaniem opisujący ewolucję funkcji stanu  $y(t) \in R^d$

$$\begin{cases} dy(t) = f(t, y(t), u(t))dt + \sigma(t, y(t), u(t))dW(t), & s \in (t_0, \infty), \\ y(t_0) = x_0, \end{cases} \quad (3)$$

gdzie  $(t_0, x_0)$  są ustalone, natomiast sterowanie  $u = u(t, \omega)$ , czyli jest w rzeczywistości funkcją dwóch argumentów, przy czym argument losowy  $\omega$  jest pomijany w zapisie. Zakładamy, że dana jest przestrzeń probabilistyczna z filtracją  $(\Omega, F, \{F_t\}_{t \geq t_0}, P)$ , na której jest zdefiniowany standardowy proces Wienera  $\Omega(\tau)$ ,  $u(t)$  jest parametrem który pełni rolę sterowania i należy do zbioru  $\mathcal{U}$  - tzw. zbioru sterowań dopuszczalnych.

Definicja. Powiemy, że  $y(t)$  jest rozwiązaniem układu (3) przy ustalonym  $u \in \mathcal{U}$  jeżeli zachodzi równanie całkowe

$$y(t) = y(t_0) + \int_{t_0}^t f(t, y(t), u(t))dt + \int_{t_0}^t \sigma(t, y(t), u(t))dW_t, \quad (4)$$

gdzie pierwsza całka jest całką Riemanna natomiast druga jest całką stochastyczną Ito, definicja i konstrukcja są klasyczne można znaleźć m.in. [Ikeda, Watanabe 1989<sup>1</sup>, Kartzas, Shreve 1991<sup>2</sup>] również w [Grygierzec 2016<sup>3</sup>]. W powyższej równości w sposób niejawni występuje oczywiście argument  $\omega \in \Omega$ , tzn. mamy do czynienia ze zmiennymi losowymi po obu stronach nierówności, i wymagamy aby zachodziła dla  $P$  prawie wszystkich (p.w.)  $\omega \in \Omega$ .

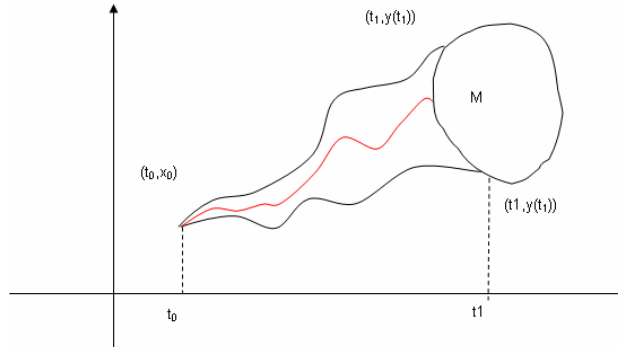
<sup>1</sup> Ikeda N., Watanabe S. (1989) Stochastic Differential Equations and Diffusion Processes, 2nd Edition. North Holland-Kodansha, Amsterdam-Tokyo.

<sup>2</sup> Kartzas I., Shreve S. E. (1991) Brownian Motion and Stochastic Calculus (Graduate Texts in Mathematics). Springer-Verlag.

<sup>3</sup> Grygierzec W. (2016) O pewnym problemie Mayera sterowania optymalnego w przypadku stochastycznym. Metody ilościowe w badaniach ekonomicznych, XVII/2, 36-45.



Rysunek 1: Problem Mayera



Źródło: opracowanie własne

Założmy że spełnione są warunki gwarantujące istnienie jednoznacznego rozwiązania  $y(t) = y(t; t_0, x_0, u)$  układu (3). Problem sterowania w przypadku Mayer'a związany jest z zadaniem zbioru  $M \subset R^d$  o regularnym brzegu  $\partial M$  oraz deterministyczną funkcją  $g(t, x)$  określona na brzegu  $\partial M$  tego zbioru, tzn.

$$g : [t_0, \infty) \times \partial M \rightarrow R, \quad (5)$$

oraz z tzw. momentem stopu  $t_1 = t_1(\omega)$ , czyli momentem Markowa, który zdefiniowany jest w sposób następujący. Jako zbiór sterowań dopuszczalnych przyjmujemy tylko takie sterowania  $u(t)$ , które w skończonym czasie przeprowadzają wektor stanu  $\psi(\tau)$  do zbioru  $M$  prawie na pewno (p.n.), oznacza to, że istnieje taki (losowy) moment stopu  $t_1 = t_1(\omega)$ , że  $y(t_1) \in M$ :

$$t_1 = \inf \{t > t_0, y(t; t_0, x_0, u) \in M \text{ p.n.}\}. \quad (6)$$

Niech będzie zadana pewna funkcja  $g$  na brzegu  $\partial M$  zbioru  $M$ :

$$g : [t_0, \infty) \times \partial M \rightarrow R. \quad (7)$$

Zdefiniujemy następujący funkcjonal kosztu:

$$J(t_0, x_0; u) = Eg(t_1, y(t_1; t_0, x_0, u)), \quad (8)$$

zbiór sterowań dopuszczalnych

$$U_{t_0, x_0} = \{u : [0, \infty) \times \Omega \rightarrow U \mid F_t - \text{adaptowane}, t_1(t_0, x_0, u(\cdot)) < \infty \text{ p.n.}\} \quad (9)$$

oraz funkcja wartości  $v$ :

$$v(t_0, x_0) = \inf_{u \in U_{t_0, x_0}} Eg(t_1, y(t_1; t_0, x_0, u)). \quad (10)$$

Zagadnienie sterowania optymalnego z powyżej zdefiniowaną funkcją wartości nazywamy problemem Mayera dla sterowania optymalnego.

Powiemy, że  $y^*$  jest trajektorią optymalną a  $(y^*, u^*)$  jest parą optymalną, jeżeli

$$v(t_0, x_0) = \inf_{u \in U_{t_0, x_0}} Eg(t_1, y(t_1; t_0, x_0, u)) = Eg(t_1, y^*(t_1; t_0, x_0, u^*)). \quad (11)$$

W teorii sterowania optymalnego rozważa się dwa zasadnicze podejścia: podejście Pontryagina, które opiera się o zasadę maksimum, dla której istnieją szerokie opracowania w literaturze m.in. [Hausman 1986<sup>4</sup>], [Grygierzec 2012<sup>5</sup>] oraz podejście Bellmana, czyli oparte o programowania dynamiczne. W tym ostatnim poszukuje się tzw. funkcji wartości, która ma tutaj kluczowe znaczenie i wykazanie pewnych jej własności jest właśnie przedmiotem niniejszego artykułu. Podobne zagadnienie w sytuacji deterministycznej jest przedstawione np. książce [Fleming Rishel 1975<sup>6</sup>]. Również monografie [Fleming, Soner 1993<sup>7</sup>] oraz [Yong, Zhou 1999<sup>8</sup>] są poświęcone problemom m.in. własności funkcji wartości. Podejście przedstawione w niniejszej pracy jest bliskie tzw. twierdzeniu weryfikacyjnemu, stanowi istotny wkład w lepsze zrozumienie i poszukiwanie rozwiązań optymalnego sterowania.

Oznaczmy przez  $Q = [0, T] \times R^d$ . Przyjmiemy następujące założenia: niech funkcje

$$\begin{aligned} f &: Q \times U \rightarrow R^d, \\ \sigma &: Q \times U \rightarrow R^d \times R^d, \end{aligned}$$

będą ciągłe oraz niech  $f(\cdot, \cdot, v)$ ,  $\sigma(\cdot, \cdot, v)$  będą klasy  $C^1(Q)$ . Zakładamy, że istnieje pewna stała  $C > 0$ , taka, że:

$$\begin{aligned} (1) & |f_t| + |f_x| + |\sigma_t| + |\sigma_x| \leq C, \\ (2) & |f(t, x, v)| + |\sigma(t, x, v)| \leq C(1 + |x| + |v|), \end{aligned}$$

gdzie  $f_t, f_x$  oznaczają odpowiednie pochodne cząstkowe  $\frac{\partial f}{\partial t}$  oraz  $\frac{\partial f}{\partial x}$ , natomiast  $|\sigma|$  jest normą operatorową  $\sigma$ , również  $\sigma_t, \sigma_x$  są pochodnymi cząstkowymi  $\sigma$  o wartościach macierzowych.

<sup>4</sup> Hausmann U. G. (1986) A Stochastic Maximum Principle for Optimal Control of Diffusions. Pitman Research Notes in Mathematics Series, 151, Longman.

<sup>5</sup> Grygierzec W. (2012) O jednolitym podejściu do rachunku wariacyjnego i sterowania optymalnego. Metody ilościowe w badaniach ekonomicznych, XIII/1.

<sup>6</sup> Fleming W. H., Rishel R. W. (1975) Deterministic and Stochastic Optimal Control. Springer-Verlag.

<sup>7</sup> Fleming W. H., Soner H. M. (1993) Controlled Markov Processes and Viscosity Solutions, Springer-Verlag.

<sup>8</sup> Yong J., Zhou X. Y. (1999) Stochastic Controls: Hamiltonian Systems and HJB Equations. Springer-Verlag.

Z ogólnej teorii stochastycznych równań różniczkowych wiadomo [Ikeda, Watanabe 1989<sup>9</sup>], że przy powyższych założeniach równanie (8) ma dokładnie jedno rozwiązanie  $y(t) \in L^2(\Omega; C(0, T; R^d))$ , gdzie

$$L^2(\Omega; C(0, T; R^d)) = \{x_t : E \sup_{t \in [0, T]} |x_t|^2 < \infty\}.$$

## WŁASNOŚCI FUNKCJI WARTOŚCI

Przy powyższych założeniach w pracy [Grygierzec 2016<sup>10</sup>] zostały udowodnione dwie własności funkcji wartości które przytoczymy tu bez dowodu:

**Twierdzenie 1**

*Proces  $v(t, y(t))$  jest podmartyngelem dla  $t \in [t_0, t_1]$  tzn.*

$$v(s, y(s)) \leq E[v(t, y(t)) | F_s] \text{ p.n.}$$

dla  $t_0 \leq s \leq t \leq t_1$ .

**Twierdzenie 2**

*$y^*$  - trajektoria optymalna  $\Rightarrow v(t, y(t)) \equiv \text{const}$  p.n.*

Wniosek który płynie z powyższych twierdzeń jest taki, że *funkcja wartości* na wszystkich trajektoriach rozwiązania układu (2) jest *podmartyngelem*, jedynie na *trajektorii optymalnej* jest stała prawie wszędzie względem miary probabilistycznej  $P$ .

W bieżące pracy wykażemy następną własność funkcji wartości, a mianowicie:

**Twierdzenie 3**

*Niech  $W(s, y)$  będzie funkcją rzeczywistą określoną na  $R^{d+1}$  taką, że zachodzi:*

- (i)  $W(s, y) = g(s, y)$  dla dowolnego  $(s, y) \in [t_0, \infty) \times \partial M$ ,
- (ii)  $W(s, y) \leq E[W(t, y(t)) | F_s]$  dla dowolnego  $t_0 \leq s \leq t_1$ ,
- (iii)  $W(t, \tilde{y}(t)) \equiv \text{const}$  p.n. dla  $t_0 \leq t \leq t_1$ ,

*Wtedy*

$$\tilde{y} \text{ jest trajektorią optymalną tzn. } \tilde{y} = y^* \text{ oraz } W(t_0, x_0) = v(t_0, x_0).$$

<sup>9</sup> Ikeda, N., Watanabe S. (1989) Stochastic Differential Equations and Diffusion Processes. 2nd Edition. North Holland-Kodansha, Amsterdam-Tokyo.

<sup>10</sup> Grygierzec W. (2016) O pewnym problemie Mayera sterowania optymalnego w przypadku stochastycznym. Metody ilościowe w badaniach ekonomicznych, XVII/2, 36-45.

Dowód. Z założeń mamy (i) i (ii) mamy

$$\begin{aligned} W(t_0, x_0) &\leq E[W(t_1, y(t_1)) | F_{t_0}] = E[W(t_1, y(t_1))] = \\ &= Eg(t_1, y(t_1)), \end{aligned}$$

natomiast z założeń (i) i (iii) mamy

$$\begin{aligned} W(t_0, x_0) &= EW(t, \tilde{y}(t)) = EW(t_1, \tilde{y}(t_1)) = \\ &= Eg(t_1, \tilde{y}(t_1)), \end{aligned}$$

z czego wynika, że

$$Eg(t_1, \tilde{y}(t_1)) \leq Eg(t_1, y(t_1))$$

zatem  $\tilde{y} = y^*$  jest trajektoria optymalną oraz

$$W(t_0, x_0) = Eg(t_1, y^*(t_1)) = v(t_0, x_0),$$

czyli  $W(t_0, x_0)$  jest funkcją wartości.

Powyższe twierdzenie pozwala na znalezienie zweryfikowanie czy jakaś funkcja jest funkcją wartości, ponieważ w ogólności funkcja ta nie jest dana w sposób jawny. Jeżeli np. odgadujemy pewną funkcję  $W(t, y)$  która to funkcja zgadza się z funkcją  $g(t, y)$  na brzegu  $\partial M$  zbioru  $M$ , i jeżeli funkcja ta jest podmartynałem na wszystkich trajektoriach a na pewnej trajektorii jest stała to funkcja ta jest funkcją wartości.

## BIBLIOGRAFIA

- Fleming W. H., Rishel R. W. (1975) Deterministic and Stochastic Optimal Control. Springer-Verlag.
- Fleming W. H., Soner H. M. (1993) Controlled Markov Processes and Viscosity Solutions. Springer-Verlag.
- Hausmann U. G. (1986) A Stochastic Maximum Principle for Optimal Control of Diffusions. Pitman Research Notes in Mathematics Series, 151, Longman.
- Grygierzec W. (2012) O jednolitym podejściu do rachunku wariacyjnego i sterowania optymalnego. Metody ilościowe w badaniach ekonomicznych, XIII/1, 118-126.
- Grygierzec W. (2016) O pewnym problemie Mayera sterowania optymalnego w przypadku stochastycznym. Metody ilościowe w badaniach ekonomicznych, XVII/2, 36-45.
- Kartzas I., Shreve S. E. (1991) Brownian Motion and Stochastic Calculus (Graduate Texts in Mathematics). Springer-Verlag.
- Ikeda N., Watanabe S. (1989) Stochastic Differential Equations and Diffusion Processes. 2nd Edition. North Holland-Kodansha, Amsterdam-Tokyo.
- Peng S. (1990) A general stochastic maximum principle for optimal control problems. SIAM Journal on Control and Optimization, 28(4), 966-979.
- Pham H. (2009) Continuous-time Stochastic Control and Optimization with Financial Applications. Springer-Verlag.
- Szafirski B. (2012) Notes from seminar. Not published.

Yong J., Zhou X. Y. (1999) Stochastic Controls: Hamiltonian Systems and HJB Equations. Springer-Verlag.

#### PROPERTIES OF VALUE FUNCTION FOR STOCHASTIC OPTIMAL CONTROL PROBLEM OF MAYER TYPE

**Abstract:** We consider stochastic optimal control problem of Mayer type. The evolution of system is described by Ito's stochastic differential equation. Such systems are sometimes called diffusion models. The cost functional relay only on terminal condition. The value function play crucial role in determining the so called feedback optimal control. In the present paper which is a continuation of previous one the authors prove some properties of value function and gives a verification criterion.

**Keywords:** stochastic optimal control, value function, Mayer problem

## ANALIZA ZASOBÓW KOMPETENCJI Z WYKORZYSTANIEM GRUPOWANIA ZMIENNYCH NIEMETRYCZNYCH

**Alicja Grześkowiak**

Wydział Zarządzania, Informatyki i Finansów  
Uniwersytet Ekonomiczny we Wrocławiu  
e-mail: alicja.grzeskowiak@ue.wroc.pl

**Streszczenie:** W artykule przedstawiono zastosowanie grupowania zmiennych niemetrycznych do badania zasobów kompetencji w Polsce. Analizę przeprowadzono na podstawie danych z ogólnopolskiego badania pt. Bilans Kapitału Ludzkiego odnoszących się zarówno do kompetencji twardych, jak i miękkich. Zastosowana procedura analityczna pozwoliła na wyróżnienie grup kompetencji podobnie ocenianych i wyznaczenie dla nich syntetycznych zmiennych – reprezentantek, które posłużyły do ewaluacji powiązań z wybranymi charakterystykami społeczno-demograficznymi.

**Słowa kluczowe:** kompetencje, zmienne niemetryczne, analiza skupień

### WSTĘP

Zasoby kompetencji pełnią bardzo istotną funkcję w gospodarce warunkując jakość kapitału ludzkiego, który stanowi jeden z ważniejszych czynników rozwoju społeczno-gospodarczego. Ich znaczenie można także rozpatrywać w kontekście działania organizacji oraz w wymiarze indywidualnym związanym z kształtowaniem kariery poszczególnych jednostek. Ze względu na kluczowe znaczenie kompetencji właściwe wydaje się prowadzenie badań nad ich stanem z wykorzystaniem stosownych procedur analitycznych. Wieloaspektowość i możliwość występowania zależności pomiędzy posiadaniem różnych umiejętności skłania do stosowania metod statystycznej analizy wielowymiarowej do oceny zasobów kompetencji, co stanowi przedmiot rozważań niniejszej pracy. Obszernym źródłem danych dotyczących poziomu kompetencji, wykorzystanych w przedstawionej analizie, są wyniki ostatniej edycji badania Bilans Kapitału Ludzkiego. Zebrane opinie na temat posiadanych kompetencji są wyrażone na skali porządkowej, co wymaga zastosowania specyficznych metod analitycznych, które

są słabiej akcentowane i poznane w przeciwieństwie do bogatego arsenału technik umożliwiających badanie zbiorów zmiennych metrycznych.

Artykuł koncentruje się wokół dwóch zasadniczych celów. Pierwszy z nich ma charakter poznawczy i dotyczy oceny relacji zachodzących pomiędzy różnymi kompetencjami, w szczególności wiąże się z wyodrębnianiem grup umiejętności podobnie ocenianych przez respondentów oraz identyfikacją ich powiązań z charakterystykami społeczno-demograficznymi ankietowanych. Drugi cel jest ukierunkowany metodycznie i odnosi się do wskazania możliwości zastosowania metody grupowania zmiennych niemetrycznych zaproponowanej w pracy [Chavent i in. 2013] do prowadzenia analiz opartych na zmiennych pochodzących ze słabych skal pomiarowych. Ma to duże znaczenie w badaniach społeczno-ekonomicznych, w których często bazuje się na wynikach badań sondażowych zdominowanych przez występowanie zmiennych nominalnych lub porządkowych. Posiadanie tego rodzaju danych nie przekreśla jednakże stosowania podejścia wielowymiarowego, a rozwój procedur analitycznych i ich implementacja w środowisku R owocuje nowymi możliwościami w tym zakresie, na co warto zwrócić uwagę przy doborze metod analizy.

## METODYKA BADANIA

Wykonane analizy opierają się na danych wtórnych pochodzących z ogólnopolskiego badania „Bilans Kapitału Ludzkiego” dotyczącego ludności, przeprowadzonego przez Polską Agencję Przedsiębiorczości oraz Uniwersytet Jagielloński w 2014 r., obejmującego swym zasięgiem 17 674 respondentów. Przed wykonaniem obliczeń usunięto nieznaczną liczbę obserwacji ze względu na występowanie braków danych.

Poziom posiadanych kompetencji był oceniany przez ankietowanych w pięciostopniowej skali porządkowej: 1 – niski, 2 – podstawowy, 3 – średni, 4 – wysoki, 5 – bardzo wysoki. Opiniowane były zarówno kompetencje twarde, jak i miękkie [zob. Górniak 2014, s. 192-193, 205-206]. Zestawienie kompetencji wziętych pod uwagę w niniejszej pracy wraz ze skrótami stosowanymi w dalszej części artykułu jest następujące [Bilans Kapitału Ludzkiego 2013]<sup>1</sup>:

- kompetencje twarde: (A) wyszukiwanie i analiza informacji oraz wyciąganie wniosków (informacje) – szybkie streszczanie dużej ilości tekstu (tekst); logiczne myślenie, analiza faktów (logika); ciągłe uczenie się nowych rzeczy (nowe); (B) obsługa, montowanie i naprawa urządzeń technicznych (urządzenia); (C) wykonywanie obliczeń (obliczenia) – wykonywanie prostych rachunków (proste); wykonywanie zaawansowanych obliczeń matematycznych

---

<sup>1</sup> Określenia kompetencji przytoczono w brzmieniu dosłownym z kwestionariusza BKL [[https://bkl.parp.gov.pl/pobierz.html/kwestionariusze\\_IV\\_2013.7z](https://bkl.parp.gov.pl/pobierz.html/kwestionariusze_IV_2013.7z)], w nawiasach podano skrócone określenia stosowane przez Autora w dalszej części opracowania.

- (zaawansowane); (D) obsługa komputera i wykorzystanie Internetu (komputer) – podstawowa znajomość pakietu typu MS Office (Office), znajomość specjalistycznych programów, umiejętność pisania programów czy tworzenia stron internetowych (specjalistyczne),
- kompetencje miękkie: (A) samoorganizacja pracy i przejawianie inicjatywy (samoorganizacja) – samodzielne podejmowanie decyzji (decyzje); przedsiębiorczość i przejawianie inicjatywy (przedsiębiorczość); kreatywność (kreatywność); odporność na stres (stres); terminowa realizacja zaplanowanych działań (terminowość); (B) kontakty z innymi ludźmi, zarówno ze współpracownikami, jak i klientami czy podopiecznymi (kontakty) – współpraca w grupie (grupa); łatwe nawiązywanie kontaktów z współpracownikami czy klientami (relacje); bycie komunikatywnym i jasne przekazywanie myśli (komunikatywność); rozwiązywanie konfliktów pomiędzy ludźmi (konflikty); (C) organizowanie i prowadzenie prac biurowych (biuro); (D) zdolności kierownicze i organizacja pracy innych (kierownicze) – koordynowanie pracy innych pracowników (koordynacja); dyscyplinowanie innych pracowników – przywoływanie ich do porządku (dyscyplinowanie); (E) dyspozycyjność (dyspozycyjność) – gotowość do częstych wyjazdów (wyjazdy); elastyczny czas pracy (elastyczność).

Rozpatrzone zostały cztery kompetencje twarde natury ogólnej oraz siedem szczegółowych. Zestaw kompetencji miękkich jest szerszy – pięć ogólnych oraz trzynaście szczegółowych. Ze względu na ich odmienny charakter analizy przeprowadzono oddzielnie dla kompetencji miękkich i twardych. Oprócz odpowiedzi dotyczących poziomu umiejętności w badaniu uwzględniono również zmienne charakteryzujące respondentów pod względem społeczno-demograficznym, tj.: miejsce zamieszkania (miasto, wieś); wiek skategoryzowany (18-24, 25-34, 35-44, 45-54, 55-59/64); płeć (kobieta, mężczyzna); wykształcenie (gimnazjalne i poniżej, zasadnicze zawodowe, średnie, wyższe); sytuację zawodową według BAEL (pracujący, bezrobotni, nieaktywni).

W przytoczonym zestawieniu zmiennych znajdują się wyłącznie zmienne o charakterze niemetrycznym, co musi mieć wpływ na wybór stosownych metod ilościowych zaprojektowanych do badania tego rodzaju danych. Duża liczba rozpatrywanych kompetencji ocenianych na skali porządkowej stanowi swoiste wyzwanie analityczne. Oczywiście możliwe jest badanie każdej kompetencji osobno, np. oceniając średnie wyniki w różnych przekrojach [zob. np. Czarnik, Turek 2015], ale ze względów merytorycznych można podejrzewać, że pomiędzy poziomem różnych kompetencji mogą istnieć powiązania. W literaturze przedmiotu można odnaleźć próby zastosowania różnych technik uwzględniających wieloaspektowe powiązania pomiędzy umiejętnościami. Badanie relacji pomiędzy kompetencjami za pomocą analizy głównych składowych dla zmiennych niemetrycznych (CatPCA) opisano w pracy [Grześkowiak 2014], a zastosowanie klasycznej analizy czynnikowej znaleźć



można w publikacji [Górniak 2014]. W artykule [Dziechciarz-Duda, Dziechciarz 2016] wykorzystano analizę korespondencji, analizę PROFIT oraz grupowanie oparte na mierze odległości GDM. W publikacji [Grześkowiak 2015] do badania związków użyto reguł asocjacyjnych, a wykorzystanie wielorakiej analizy czynnikowej uwzględniającej grupy wieku przedstawiono w pracy [Grześkowiak 2016]. Rezultaty wymienionych prac wskazują na istnienie powiązań pomiędzy deklarowanym poziomem różnych umiejętności. Taka struktura danych skłania do poszukiwań pewnych uogólnień, które w sposób syntetyczny pozwalałyby opisywać złożone zasoby kompetencyjne.

W niniejszej pracy zasugerowano zastosowanie procedury grupowania zmiennych zaproponowanej i omówionej w [Chavent i in. 2013] do identyfikacji powiązań pomiędzy kompetencjami i ich odzwierciedlenia za pomocą nowych reprezentantek. Na korzyści płynące z grupowania zmiennych zwraca się uwagę w [Lasek, Pęczkowski 2010] wskazując na likwidację problemu współliniowości i redundancji informacji, zwiększenie przejrzystości i czytelności związków między zmiennymi oraz umożliwianie budowy modeli o mniejszej złożoności.

W literaturze traktującej o grupowaniu zmiennych spotkać można dwa podejścia: najczęstszym jest zastosowanie takich samych algorytmów, jak przy klasyfikacji obiektów, opisanych np. w pracach [Kaufman, Rousseeuw 2009, Everitt i in. 2011, Kassambara 2017]. Podejście drugie to wykorzystanie procedur opracowanych wyłącznie w tym celu, ale należy zaznaczyć, że metod poświęconych jedynie grupowaniu zmiennych jest stosunkowo niewiele. Jak wskazano w opracowaniu [Chavent i in. 2013] do najważniejszych metod dla zmiennych metrycznych należy zaliczyć: procedurę VARCLUS dostępną w oprogramowaniu SAS, metodę grupowania wokół zmiennych ukrytych [Vigneau, Qannari 2003, Vigneau i in. 2015], diametrical clustering [Dhillon i in. 2003], ujęcie nieparametryczne [Palla i in. 2012] oraz podejście bazujące na analizie kanonicznej [Bühlmann i in. 2013]. Chavent i in. [2013] przedstawiają uniwersalne rozwiązanie problemu grupowania zmiennych, które zostanie wykorzystane w niniejszym artykule. Jest ono szczególnie interesujące dla badaczy stykających się z analizą danych niemetrycznych lub danych o mieszanym charakterze (metrycznych i niemetrycznych). Autorzy koncepcji postulują tworzenie zmiennych syntetycznych  $c_k$  reprezentujących wyodrębniane skupienia  $C_k$  o postaci [Chavent i in. 2013]:

$$c_k = \arg \max_{u \in \mathbb{R}^n} \left\{ \sum_{x_j \in C_k} r_{u|x_j}^2 + \sum_{y_j \in C_k} \eta_{u|y_j}^2 \right\} \quad (1)$$

gdzie:  $x_i$  – zmienne metryczne,  $y_i$  – zmienne niemetryczne,  $r_{u|x_j}$  – współczynnik korelacji Pearsona,  $\eta_{u|y_j}^2$  – stosunek korelacyjny,  $u$  – wartości, dla których jest poszukiwane maksimum.

Zmienne syntetyczne  $c_k$  są otrzymywane jako pierwsze główne składowe otrzymane z procedury PCAMIX, tj. analizy głównych składowych dla mieszaniki zmiennych metrycznych i niemetrycznych [Kiers 1991]. Algorytmy grupowania są

tak skonstruowane by zmaksymalizować kryterium homogeniczności, które dla danego skupienia  $C_k$  definiowane jest jako [Chavent i in. 2013]:

$$H(C_k) = \sum_{x_j \in C_k} r_{u,x_j}^2 + \sum_{y_j \in C_k} \eta_{u|y_j}^2, \quad (2)$$

natomiast dla całego podziału jako suma miar homogeniczności (2) wyznaczonych dla utworzonych klas. Dokładny opis algorytmów wraz z charakterystyką pakietu `ClustOfVar` programu R pozwalającego na zastosowanie procedur jest przedstawiony w pracy [Chavent i in. 2013]. Wartościowym aspektem, na który zwracają uwagę autorzy idei, jest otrzymanie ilościowych zmiennych syntetycznych reprezentujących poszczególne skupienia. Pakiet `ClustOfVar` daje możliwość wyboru metody grupowania spośród hierarchicznej procedury aglomeracyjnej oraz metody k-średnich. W niniejszej pracy zastosowano pierwsze z wymienionych podejść. Ze stosowaniem hierarchicznych procedur aglomeracyjnych wiąże się problem ustalania ostatecznego podziału na klasy, gdyż metoda w sama w sobie nie daje odpowiedzi na pytanie, ile należy wyodrębnić grup. Twórcy pakietu `ClustOfVar` proponują ocenę stabilności podziału za pomocą metody bazującej na średnim skorygowanym indeksie Randa.

Porównanie dwóch różnych wyników podziału  $P$  i  $P'$  można oprzeć na spostrzeżeniu, że istnieją cztery możliwości: (a) dwa elementy znajdują się w tej samej grupie zarówno w podziale  $P$ , jak i  $P'$ , (b) w podziale  $P$  są w tej samej grupie, a w podziale  $P'$  nie, (c) w podziale  $P'$  są w tej samej grupie, a w podziale  $P$  nie, (d) znajdują się w różnych grupach w obu podziałach [zob. Wagner, Wagner 2007]. Indeks Randa [Rand 1971] bazuje na porównaniu liczby kombinacji zgodnych do wszystkich wariantów:

$$R = \frac{a+d}{a+b+c+d}, \quad (3)$$

Ze względu na fakt, że wartość oczekiwana indeksu Randa dwóch losowych podziałów nie jest stała, Hubert i Arabie zaproponowali jego modyfikację określaną skorygowanym indeksem Randa [Hubert, Arabie 1985], który można wyrazić formułą [Yeung, Ruzzo 2001]:

$$R_{sk} = \frac{\sum_{i,j} \binom{n_{ij}}{2} - [\sum_i \binom{n_{i.}}{2} \sum_j \binom{n_{.j}}{2}] / \binom{n}{2}}{0,5[\sum_i \binom{n_{i.}}{2} + \sum_j \binom{n_{.j}}{2}] - [\sum_i \binom{n_{i.}}{2} \sum_j \binom{n_{.j}}{2}] / \binom{n}{2}} \quad (4)$$

gdzie  $n_{ij}$ ,  $n_{i.}$ ,  $n_{.j}$  – odpowiednio liczebności komórek oraz liczebności brzegowe tabeli kontyngencji odzwierciedlającej przynależność do grup otrzymanych w dwóch podziałach  $P$  i  $P'$ .

Pakiet `ClustOfVar` oferuje podejście bootstrapowe do obliczania zmodyfikowanego indeksu Randa umożliwiającego ocenę stabilności podziałów. Średni skorygowany indeks Randa według formuły (4) obliczony na podstawie pięćdziesięciu prób typu bootstrapowego stanowił przesłankę do wyboru ostatecznej liczby klas.

Otrzymane grupy zmiennych skupiające kompetencje o zbliżonych ocenach są reprezentowane przez ilościowe zmienne syntetyczne postaci (1), które stanowią

podstawę dalszych analiz wiążących ocenę umiejętności z cechami demograficzno-społecznymi. W celu oceny siły powiązań posłużono się współczynnikiem eta, adekwatnym do pomiaru stopnia związku pomiędzy zmienną metryczną a niemetryczną, który można wyrazić jako [zob. Malarska 2005, s.89]:

$$\eta_z = \frac{s(\bar{z}_j)}{s(z)}, \quad s^2(\bar{z}_j) = \frac{1}{n} \sum_{j=1}^k (\bar{z}_j - \bar{z})^2 n_{.j}, \quad (5)$$

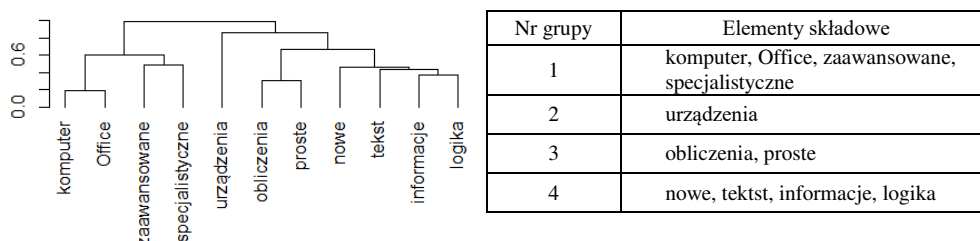
gdzie:  $s(z)$  – odchylenie standardowe rozpatrywanej zmiennej metrycznej,  $k$  – liczba kategorii zmiennej niemetrycznej,  $\bar{z}_j$  – średnia zmiennej metrycznej dla danej kategorii zmiennej niemetrycznej. Współczynnik eta jest miarą unormowaną w przedziale  $\langle 0,1 \rangle$ , co pozwala na ocenę i porównywanie siły zależności.

## WYNIKI BADAŃ EMPIRYCZNYCH

W niniejszym rozdziale przedstawiono wyniki analizy zasobów kompetencji w podziale na kompetencje twarde i miękkie. W odniesieniu do każdego z obszarów przeprowadzono grupowanie zmiennych niemetrycznych i wyodrębniono ich skupienia, których liczbę określono na podstawie kształtowania się skorygowanego indeksu Randa. Dla otrzymanych w ten sposób klas obliczono zmienne syntetyczne je reprezentujące, które następnie wykorzystano do oceny powiązań ze zmiennymi demograficznymi.

Na rysunku 1 zilustrowano wyniki procedury grupowania zmiennych niemetrycznych odzwierciedlających kompetencje twarde.

Rysunek 1. Wyniki hierarchicznej procedury aglomeracyjnej dla kompetencji twardych



Źródło: opracowanie własne

Otrzymano cztery klasy, w tym jedną jednoelementową – jest to skupienie nr 2 zawierające wyłącznie umiejętności w zakresie obsługi urządzeń technicznych, co oznacza, że poziom tej kompetencji charakteryzuje się dużą odrębnością i nie wiąże się z deklarowaniem innych umiejętności. Skupienie nr 3 łączy kompetencje matematyczne jako kategorię ogólną z wykonywaniem prostych rachunków. Warto zauważyć, że wykonywanie zaawansowanych obliczeń nie zostało przyłączone do tej klasy, lecz stanowi element skupienia nr 1 obejmującego ponadto wszystkie rozpatrywane kompetencje informatyczne, zarówno na poziomie podstawowym,

jak i specjalistycznym. Ostatnia z wyodrębnionych grup zawiera umiejętności związane z wyszukiwaniem i analizą informacji oraz uczeniem się nowych rzeczy.

Powiązania zmiennych syntetycznych reprezentujących poszczególne skupienia z charakterystykami demograficzno-społecznymi odzwierciedlają współczynniki eta zawarte w tabeli 1.

Tabela 1. Wartości współczynników eta pomiędzy zmiennymi syntetycznymi dla kompetencji twardych i charakterystykami demograficzno-społecznymi

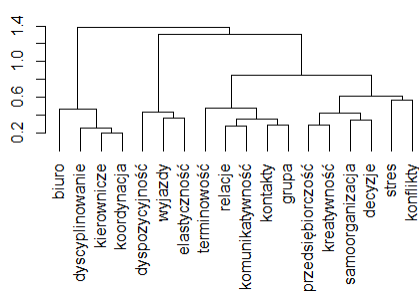
Nr grupy	Płeć	Wiek	Miejsce zamieszkania	Wykształcenie	Sytuacja zawodowa
1	0,053	0,423	0,166	0,497	0,170
2	0,294	0,048	0,009	0,091	0,152
3	0,063	0,163	0,102	0,407	0,183
4	0,107	0,241	0,147	0,501	0,211

Źródło: opracowanie własne

Z wyjątkiem skupienia nr 2 (obsługa urzędów) najwyższe wartości miary otrzymano dla poziomu wykształcenia. Można więc stwierdzić, że ten czynnik w największym stopniu jest związany z oceną kompetencji twardych, w szczególności wysoko z umiejętnościami dotyczącymi analizy informacji oraz kompetencjami informatycznymi w powiązaniu z wykonywaniem zaawansowanych obliczeń. Należy również podkreślić wysoki stopień powiązania zmiennej syntetycznej dla skupienia nr 1 z wiekiem ankietowanych, co wiąże się z lepszą znajomością nowoczesnych technologii przez ludzi młodych. Z kolei umiejętność obsługi urzędów pozostaje w zależności z płcią respondentów, co wynika z lepszej samooceny mężczyzn w tym zakresie.

Analogiczna procedura przeprowadzona dla zestawu kompetencji miękkich wskazała na istnienie sześciu grup (rysunek 2).

Rysunek 2. Wyniki hierarchicznej procedury aglomeracyjnej dla kompetencji miękkich



Nr grupy	Elementy składowe
1	biuro, dyscyplinowanie, kierownicze, koordynacja
2	dyspozycyjność, wyjazdy, elastyczność
3	terminowość, relacje, komunikatywność, kontakty, grupa
4	przedsiębiorczość, kreatywność, samoorganizacja, decyzje
5	stres
6	konflikty

Źródło: opracowanie własne

Istnieją dwa rodzaje umiejętności, które stanowią odrębne klasy, mimo iż w badaniu Bilans Kapitału Ludzkiego były rozpatrywane jako kompetencje

szczegółowe stanowiące składowe kompetencji o charakterze ogólnym. Należą do nich odporność na stres (skupienie nr 5) oraz radzenie sobie w sytuacjach konfliktowych (skupienie nr 6). Wyniki aglomeracji wskazują na podobieństwo w ocenie zdolności kierowniczych oraz organizowania i prowadzenia prac biurowych (skupienie nr 1). W skupieniu nr 2 znalazły się wszystkie zmienne reprezentujące kwestie związane z dyspozycyjnością. Grupa nr 3 składa się z umiejętności interpersonalnych oraz przestrzegania terminów wykonywania zadań. Skupienie 4 obejmuje kompetencje związane z operatywnością i inwencją twórczą.

W tabeli 2 ujęto wartości współczynników eta pomiędzy zmiennymi syntetycznymi reprezentującymi skupienia a zmiennymi demograficzno-społecznymi.

Tabela 2. Wartości współczynników eta pomiędzy zmiennymi syntetycznymi dla kompetencji miękkich i charakterystykami demograficzno-społecznymi

Nr grupy	Płeć	Wiek	Miejsce zamieszkania	Wykształcenie	Sytuacja zawodowa
1	0,071	0,157	0,132	0,444	0,233
2	0,136	0,165	0,072	0,248	0,315
3	0,077	0,156	0,072	0,352	0,237
4	0,052	0,208	0,089	0,420	0,274
5	0,018	0,090	0,015	0,137	0,130
6	0,045	0,074	0,066	0,237	0,114

Źródło: opracowanie własne

Najwyższe wartości współczynników wystąpiły w przypadku poziomu wykształcenia, choć zależności te są słabsze niż dla kompetencji twardych. W najwyższym stopniu z wykształceniem korelują umiejętności kierownicze i organizowania prac biurowych (skupienie 1) oraz kompetencje ze sfery przedsiębiorczości i samoorganizacji pracy (skupienie 4). Kompetencje tworzące odrębne skupienia (nr 5 i nr 6) nie są mocno związane z żadną z rozpatrywanych cech demograficzno-społecznych. Warto zauważyć, że ocena umiejętności w bardzo niewielkim stopniu jest powiązana z płcią i miejscem zamieszkania respondentów. Z kolei kategoria wieku wykazuje najsilniejszą relację ze skupieniem nr 4, którego domena to przedsiębiorczość.

## PODSUMOWANIE

Zasadniczym wkładem artykułu jest zastosowanie rzadko używanego podejścia w analizie danych ze słabych skal pomiarowych – procedury grupowania zmiennych niemetrycznych w celu wyodrębnienia klas cech podobnych i wyznaczenia odpowiadających im zmiennych syntetycznych. Otrzymane w ten sposób reprezentantki skupień posłużyły do identyfikacji powiązań z charakterystykami

społeczno-demograficznymi respondentów, które okazały się być najsilniejsze w przypadku wykształcenia i wieku.

Należy zaznaczyć, że aplikacje metod klasyfikacji obejmują najczęściej wyodrębnianie skupień obiektów, a nie zmiennych. Natomiast przedstawione w artykule odmienne podejście, tj. grupowanie zmiennych pozwoliło na redukcję wymiarów i przedstawienie struktury ocen kompetencji w uogólniony sposób. Procedura ujawniła istnienie kompetencji, które pozostają w izolacji od pozostałych, oraz umiejętności, które pozostają ze sobą w bliskiej relacji.

Efekty analiz zaprezentowanych w pracy wskazują, że warto sięgać po specyficzne metody analityczne adekwatne dla zmiennych o charakterze niemetrycznym. Ze względu na ograniczone ramy niniejszej pracy analizy zostały przeprowadzone wyłącznie za pomocą hierarchicznej procedury aglomeracyjnej. Wykorzystana metoda daje natomiast podstawy do pogłębienia rozważań o inne metody analiz wielowymiarowych, np. grupowanie metodą k-średnich, co wskazuje na potencjalne dalsze kierunki badań. Kolejnym obszarem zainteresowania może być wykorzystanie omówionych metod w celu oceny zmian zachodzących w zasobach kompetencji w wymiarze czasowym.

## BIBLIOGRAFIA

- Bilans Kapitału Ludzkiego w Polsce. Badania ludności (2013) Polska Agencja Rozwoju Przedsiębiorczości, [https://bkl.parp.gov.pl/pobierz.html/kwestionariusze\\_IV\\_2013.7z](https://bkl.parp.gov.pl/pobierz.html/kwestionariusze_IV_2013.7z).
- Bühlmann P., Rütimann P., van de Geer S., Zhang C. H. (2013) Correlated Variables in Regression: Clustering and Sparse Estimation. *Journal of Statistical Planning and Inference*, 143(11), 1835-1858.
- Chavent M., Kuentz V., Liqueur B., Saracco L. (2013) ClustOfVar: An R Package for the Clustering of Variables. *Journal of Statistical Software*, 50(13), 1-16.
- Czarnik Sz., Turek K. (2015) Polski rynek pracy – aktywność zawodowa i struktura wykształcenia. PARP, Warszawa.
- Dhillon I. S., Marcotte E. M., Roshan U. (2003) Diametrical Clustering for Identifying anti-Correlated Gene Clusters. *Bioinformatics*, 19(13), 1612-1619.
- Dziechciarz-Duda M., Dziechciarz J. (2016) The Identification of Training Needs for Human Capital Quality Improvement in Poland – a Statistical Approach. *Statistics in Transition New Series*, 17(4), 723-736.
- Everitt B. S., Landa S., Leese M., Stahl D. (2011) *Cluster Analysis*. Wiley, Chichester.
- Górniak J. (red.) (2014) *Kompetencje Polaków a potrzeby polskiej gospodarki. Raport podsumowujący IV edycję badań BKL z 2013 roku*. Polska Agencja Rozwoju Przedsiębiorczości, Warszawa.
- Grześkowiak A. (2014) Deklarowane umiejętności zawodowe Polaków – analiza na podstawie danych sondażowych o charakterze niemetrycznym. *Acta Universitatis Nicolai Copernici. Nauki Humanistyczno-Społeczne. Ekonomia*, 45 (2), 2014, 161-171.
- Grześkowiak A. (2015) Competences of Adult Poles - Evaluation of the Age as a Differentiating Factor. [w:] Rotschedl J., Cermakova K. (red.) *Proceedings of the 15th*

- International Academic Conference. International Institute of Social and Economic Sciences, Prague, 381-391.
- Grzeškowiak A. (2016) Wielowymiarowa analiza kompetencji zawodowych według grup wieku ludności. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 427, 122-130.
- Hubert L., Arabie P. (1985) Comparing Partitions. *Journal of Classification*, 2(1), 193-218.
- Kassambara A. (2017) *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning*. STHDA.
- Kaufman L., Rousseeuw P. J. (2009) *Finding Groups in Data: an Introduction to Cluster Analysis*. John Wiley & Sons, Hoboken.
- Kiers H. A. (1991) Simple Structure in Component Analysis Techniques for Mixtures of Qualitative and Quantitative Variables. *Psychometrika*, 56(2), 197-212.
- Lasek M., Pęczkowski M. (2010) Grupowanie zmiennych w procesach eksploracji danych (Data Mining). *Ekonomia i Zarządzanie*, 2(1), 83-94.
- Malarska A. (2005) *Statystyczna analiza danych wspomagana programem SPSS*. SPSS Polska, Kraków.
- Palla K., Ghahramani Z., Knowles D. A. (2012) A Nonparametric Variable Clustering Model. [w:] *Advances in Neural Information Processing Systems*, 2987-2995.
- Rand W. M. (1971) Objective Criteria for the Evaluation of Clustering Methods. *Journal of the American Statistical Association*, 66(336), 846-850.
- Vigneau E., Chen M., Qannari E. M. (2015) ClustVarLV: An R Package for the Clustering of Variables Around Latent Variables. *The R Journal*, 7 (2), 134-148.
- Vigneau E., Qannari E. M. (2003) Clustering of Variables Around Latent Components. *Communications in Statistics-Simulation and Computation*, 32(4), 1131-1150.
- Wagner S., Wagner D. (2007) *Comparing Clusterings: an Overview*. Karlsruhe: Universität Karlsruhe, Fakultät für Informatik. [http://www.cs.ucsb.edu/~veronika/MAE/wagner07\\_comparingclusterings.pdf](http://www.cs.ucsb.edu/~veronika/MAE/wagner07_comparingclusterings.pdf).
- Yeung K. Y., Ruzzo W. L. (2001) Details of the Adjusted Rand Index and Clustering Algorithms, Supplement to the Paper An Empirical Study on Principal Component Analysis for Clustering Gene Expression Data. *Bioinformatics*, 17(9), 763-774.

#### **ANALYSIS OF COMPETENCES RESOURCES USING THE CLUSTERING METHOD FOR NONMETRIC VARIABLES**

**Abstract:** The article presents the application of a nonmetric variables clustering method to a study of competences resources in Poland. The analysis is based on results from a nationwide survey on Human Capital in Poland, including data both on hard and soft skills. The analytical procedure allowed for grouping competences assessed similarly into clusters and assigning to them synthetic variables – representatives, which were used to evaluate the relationships with chosen socio-demographic characteristics.

**Keywords:** competences, nonmetric variables, cluster analysis

## ZANURZANIE OBSERWACJI W PRÓBIE W OCENIE ZRÓŻNICOWANIA PRZESTĘPCZOŚCI PRZECIWKO MIENIU ORAZ STOPY BEZROBOCIA W POLSCE

**Małgorzata Kobylińska**

Wydział Nauk Ekonomicznych  
Uniwersytet Warmińsko-Mazurski w Olsztynie  
e-mail:angosiak@poczta.onet.pl

**Streszczenie:** W artykule przedstawiono wykorzystanie koncepcji zanurzania obserwacji w próbie w analizie danych dotyczących liczby przestępstw przeciwko mieniu oraz stopy bezrobocia. Dokonano rangowania województw względem oddalenia od centralnego skupienia zbiorów danych. Wykresy oparte na miarach zanurzania obserwacji w próbie pozwoliły na zobrazowanie pewnych własności tych danych.

**Słowa kluczowe:** zanurzania obserwacji w próbie, miara zanurzania Tukeya, miara zanurzania simpleksowego Liu, kontury zanurzania

### WSTĘP

We współczesnym świecie przestępczość jest jednym z głównych problemów społecznych. Kluczowe staje się wykrywanie przestępczości oraz jej zapobieganie. Zauważyć można wzrost zainteresowania tym zagadnieniem przez społeczeństwo, organy ścigania jak i ośrodki badań naukowych [Sztudynger i Sztudynger 2003].

Przestępczość ma istotny wpływ na codzienne życie lokalnych społeczności oraz całego społeczeństwa. Zasadniczo wpływa na poczucie bezpieczeństwa, które jest podstawowym wyznacznikiem jakości życia.

Przestępstwa przeciwko mieniu pojmowane są jako napady na własność lub inne prawa majątkowe m.in. kradzieże, włamania, paserstwa, rozboje lub niszczenie mienia. Istotą tych przestępstw jest spowodowanie poszkodowanemu straty majątkowej [Zawłocki 2011]. Na przestępczość przeciwko mieniu istotny wpływ ma wiele czynników mających charakter społeczno-ekonomiczny.



Bezrobociu przypisuje się ważną rolę na kształtowanie jego poziomu [patrz np. Szczepaniec 2011, Kędziółka 2014].

Zachowania przestępcze przybierają coraz nowsze, bardziej skomplikowane i szkodliwe formy. Istotną rolę w walce z przestępczością odgrywają badania empiryczne, które powinny być wykorzystywane przez instytucje i organizacje, których celem jest jej zwalczanie i zapobieganie. Badania nad przestępczością, mają na celu określenie czynników na nią wpływających oraz możliwości jej zapobiegania [Hołyst 2016].

Celem artykułu jest zaprezentowanie przykładu empirycznego, który ilustruje zastosowanie miar i konturów zanurzania obserwacji w próbie w analizie danych wielowymiarowych. W tym celu wykorzystano dane liczbowe dotyczące liczby przestępstw przeciwko mieniu oraz stopy bezrobocia w województwach Polski. Przykłady wykorzystania zanurzania w analizie danych wielowymiarowych zamieszczone zostały między innymi w pracy Kosiorowskiego [2012].

## TEORETYCZNE PODSTAWY ZANURZANIA OBSERWACJI W PRÓBIE

Pojęcie zanurzania obserwacji w próbie zostało wprowadzone przez Tukeya [1975]. Stało się ono inspiracją do napisania wielu prac z tego zakresu. Zanurzanie obserwacji w próbie umożliwia porządkowanie punktów w wielowymiarowej przestrzeni cech względem centrum rozkładu lub może być wykorzystywane jako narzędzie do wizualizacji danych wielowymiarowych [Donoho i Gasko 1992]. Charakterystyka kryteriów wyznaczania miary zanurzania obserwacji w próbie, jej własności oraz przykłady jej zastosowania zamieszczone zostały między innymi w pracach Liu i in. [1999], Zuo i Serfling [2000], Mosler [2002].

Niech  $P_n^p$  określa próbę  $p$ -wymiarową o liczebności  $n$ , która pochodzi z pewnego  $p$ -wymiarowego rozkładu określonego dystrybuantą  $F_p$ . Każdy punkt próby  $x_{ij}$  rozumiany jest jako wartość  $j$ -tej zmiennej, która zaobserwowana została dla  $i$ -tego obiektu. W artykule miary zanurzania zostały wyznaczone dla każdego punktu  $\theta$  (województwa) należącego do danych zbiorów danych. W szczególności punkt  $\theta$  może być dowolnym punktem z przestrzeni rzeczywistej  $R^p$ .

W literaturze przedmiotu zdefiniowane zostały różne kryteria wyznaczania miary zanurzania obserwacji w próbie. Przedstawione zostaną definicje zanurzania obserwacji w próbie, które wykorzystano w niniejszej pracy.

Miarą zanurzania sympleksowego Liu [Liu 1990] ( $Lzan_p$ ) punktu  $\theta$  w próbie  $P_n^p$  nazywamy funkcję

$$Lzan_p(\theta, P_n^p) = \left( \binom{n}{p+1} \right)^{-1} \sum_{1 \leq i_1 < i_2 < \dots < i_{p+1} \leq n} I[\theta \in \Delta(x_{i_1}, x_{i_2}, \dots, x_{i_{p+1}})], \quad (1)$$

gdzie  $\Delta(x_{i_1}, x_{i_2}, \dots, x_{i_{p+1}})$  są domkniętymi sympleksami wyznaczonymi przez  $p+1$  punktów próby, natomiast  $I[A]$  jest funkcją wskaźnikową przyjmującą wartość 1, jeżeli punkt  $\theta$  należy do sympleksu, lub wartość 0 w przeciwnym przypadku.

Minimalną liczbę punktów należących do zamkniętej półprzestrzeni, której linia rozdzielająca przechodzi przez punkt  $\theta$  nazywamy półprzestrzenią zanurzenia  $H_p$  punktu  $\theta \in R^p$  próby  $P_n^p$ .

Miarą zanurzenia Tukeya [Tukey 1975] ( $Tzan_p$ ) punktu  $\theta$  w próbie  $P_n^p$  nazywamy funkcję

$$Tzan_p(\theta, P_n^p) = \frac{1}{n} \inf_{H_p} \{ \theta \in H_p \}, \quad (2)$$

gdzie  $H_p$  jest domkniętą półprzestrzenią w  $R^p$  i  $\theta \in H_p$ .

Wartości miary zanurzenia obserwacji w próbie są nieujemne i ograniczone, należą do przedziału od 0 do 1. Pozwalają na uporządkowanie obserwacji względem centrum próby. Wyższe wartości miary zanurzenia odpowiadają obserwacjom położonym bardziej centralnie w danej „chmurze danych” niższe obserwacjom, które są najbardziej oddalone od tego centrum, ze względu na wysokie lub niskie wartości cech diagnostycznych.

Konturem zanurzenia stopnia  $k$  nazywamy zbiór

$Con_k = \{ \theta : zan_p(\theta, P_n^p) = k \}$  dla  $k=1, 2, \dots, [n/2]$ , gdzie  $[n/2]$  jest częścią całkowitą liczby  $n/2$ , natomiast  $k$  minimalną liczbę punktów należących do półprzestrzeni  $H_p$ .

Wierzchołki konturów zanurzenia wyznaczone są przez punkty przecięcia prostych, będących krawędziami półpłaszczyzn o określonej wartości zanurzenia. Wierzchołki zewnętrznego wielokąta (powłoki wypukłej) zostały wyznaczone przez obserwacje, dla których zanurzenie Tukeya wynosi zero. Szczegółowo konstrukcja konturów zanurzenia w przypadku dwuwymiarowym została przedstawiona w pracy Rousseeuwa i Ruts [1996]. Metody wyznaczania miar zanurzenia w przypadku dwuwymiarowym zaprezentowano również przez Wagnera i Kobylińską [2002].

Wykres pudełkowy dla danych dwuwymiarowych został zaproponowany przez Rousseeuwa i in. [1999]. Do jego konstrukcji wykorzystywane są półpłaszczyzny zanurzenia Tukeya. Głównym elementem wykresu jest wielokąt zawierający 50% obserwacji, którym odpowiadają najwyższe wartości zanurzenia Tukeya. Wewnątrz tego wielokąta zaznaczono medianę Tukeya, punkt o maksymalnej wartości zanurzenia.

Kontury zanurzenia mogą być wykorzystywane do wizualizacji danych wielowymiarowych. Umożliwiają zobrazowanie pewnych własności zbiorów

danych między innymi punktu centralnego skupienia (mediany Tukeya), wielkości rozproszenia lub symetrii analizowanych zbiorów danych. Kształt i orientacja konturów zanurzania wskazują na siłę i kierunek zależności pomiędzy badanymi zmiennymi [Chabana i Ouarda 2011].

Wybrane kryteria wyznaczania miary zanurzania obserwacji w próbie zostały porównane na materiale empirycznym w pracy Kobylińskiej [2006].

## WYNIKI BADAŃ

Do analizy wykorzystano dane zawarte w Banku Danych Lokalnych GUS z lat 2005, 2010 i 2015. Do obliczenia wartości miar zanurzania (tabela 3) oraz sporządzenia wykresów opartych na zanurzeniu obserwacji w próbie (rysunek 1 i 2), wykorzystane zostały pakiety środowiska R: „depth” autorstwa Maxime Genest, Jean-Claude Masse, Jean-Francois Plante<sup>1</sup>, „depthproc” autorstwa Daniela Kosiorowskiego, Mateusza Bociana, Anny Wegrzynowskiej i Zygmunta Zawadzkiego<sup>2</sup> oraz „mrfDepth” autorstwa Pietera Segarta, Mia Hubert, Petera Rousseuwa, Jakoba Raymaekersa i Kaveha Vakilia<sup>3</sup>. Wstępna analiza statystyczna została przeprowadzona z wykorzystaniem pakietu statystycznego Statistica PL.

Do analizy przyjęto następujące zmienne diagnostyczne:

Y - liczba przestępstw przeciwko mieniu stwierdzonych przez Policję w zakończonych postępowaniach przygotowawczych, przypadająca na 1000 mieszkańców,

X - stopa bezrobocia ogółem (%)

Wartości prezentowanych zmiennych zestawiono w tabeli 1.

Zaobserwować można w danych latach tendencję spadkową liczby przestępstw przeciwko mieniu w województwach (tabela 1). Największa liczba przestępstw w przeliczeniu na 1000 mieszkańców występowała w województwach dolnośląskim oraz pomorskim. W województwach tych w latach 2010 i 2015 zanotowano wartość stopy bezrobocia poniżej przeciętnej w kraju. Najwyższą wartością tej cechy w danych latach charakteryzowało się województwo warmińsko-mazurskie. Najniższy wskaźnik przestępstw przeciwko mieniu zanotowano w województwie podkarpackim. Jego wartość w 2005 roku była ponad dwukrotnie niższa w porównaniu ze średnią w kraju. W tym roku przeciętna liczba przestępstw przeciwko mieniu na 1000 mieszkańców w Polsce była wyższa od odnotowanej w 2015 roku o 45,34%. Największy spadek stopy bezrobocia w 2015 roku w porównaniu z rokiem 2005 (o ponad 10 p.p.) odnotowano w województwach dolnośląskim, lubuskim, pomorskim, warmińsko-mazurskim oraz

---

<sup>1</sup> <https://cran.r-project.org/web/packages/depth/index.html> [dostęp: 15.06.2017]

<sup>2</sup> <https://cran.r-project.org/web/packages/DepthProc/index.html> [dostęp: 15.06.2017]

<sup>3</sup> <https://cran.r-project.org/web/packages/mrfDepth/index.html> [dostęp: 15.06.2017]

zachodniopomorskim. W roku 2015 w czterech województwach (kujawsko-pomorskim, małopolskim, pomorskim, zachodniopomorskim) liczba przestępstw przeciwko mieniu przypadająca na 1000 mieszkańców zmniejszyła się o ponad 50% w porównaniu z rokiem 2005.

Tabela 1. Wartości liczbowe zmiennych

Lp.	Województwo	2005		2010		2015	
		Y	X	Y	X	Y	X
1	Dolnośląskie	24,56	20,60	23,00	13,10	17,01	8,50
2	Kujawsko-pomorskie	22,41	22,30	15,16	17,00	10,12	13,20
3	Lubelskie	14,00	17,00	9,82	13,10	7,16	11,70
4	Lubuskie	19,60	23,00	17,31	15,50	13,75	10,50
5	Łódzkie	21,40	17,90	14,64	12,20	11,15	10,30
6	Małopolskie	21,09	13,80	15,19	10,40	10,34	8,30
7	Mazowieckie	24,20	13,80	14,97	9,70	12,36	8,30
8	Opolskie	17,93	18,70	15,65	13,60	10,74	10,10
9	Podkarpackie	9,98	18,50	7,91	15,40	6,65	13,20
10	Podlaskie	12,64	15,60	9,89	13,80	7,32	11,80
11	Pomorskie	28,76	19,20	21,31	12,30	12,62	8,90
12	Śląskie	26,94	15,50	21,22	10,00	14,15	8,20
13	Świętokrzyskie	16,88	20,60	10,44	15,20	9,01	12,50
14	Warmińsko-mazurskie	17,29	27,20	13,84	20,00	10,52	16,20
15	Wielkopolskie	18,92	14,60	13,79	9,20	10,81	6,10
16	Zachodniopomorskie	24,20	25,60	18,21	17,80	11,63	13,10

Źródło: opracowanie własne na podstawie danych GUS

Współczynniki zmienności (tabela 2) świadczą o najsilniejszym przestrzennym zróżnicowaniu badanych zmiennych w 2005 roku. Największe zakresy zmienności analizowanych zmiennych odnotowano w 2005 roku, 18,78 przestępstw przeciwko mieniu na 1000 mieszkańców oraz 13,4 % w przypadku stopy bezrobocia.

Tabela 2. Charakterystyki liczbowe badanych zmiennych

Charakterystyki	2005		2010		2015	
	Y	X	Y	X	Y	X
Minimum	9,98	13,80	7,91	9,20	6,65	6,10
Maksimum	28,76	27,20	23,00	20,00	17,01	16,20
Średnia	20,05	18,99	15,15	13,64	10,96	10,68
Współczynnik zmienności	28,76	27,20	23,00	20,00	17,01	16,20
Odchylenie standardowe	5,18	4,05	4,36	3,07	2,72	2,60

Źródło: opracowanie własne

W tabeli 3 zamieszczone zostały wartości miar zanurzania Tukeya oraz sympleksowego Liu wyznaczone według wzorów 1 i 2. Obok w kolumnach podano numery województw (W) zgodnie z tabelą 1. Zauważyć można, że niektórym województwom odpowiadają jednocześnie wartości miary zanurzania Tukeya

równe zero (województwa te są wierzchołkami powłok wypukłych rozpatrywanych zbiorów danych) oraz najniższe wartości miary zanurzania Liu. Wśród jednostek będących wierzchołkami powłoki wypukłej mogą występować obserwacje nietypowe. Województwa, którym odpowiadają najwyższe wartości miary zanurzania położone są najbardziej centralnie w zbiorach danych i wyznaczają mediany dwuwymiarowe tych zbiorów. W kolejnych latach mediany zostały wyznaczone przez województwa łódzkie oraz opolskie. Kwartył pierwszy i kwartył trzeci wyznaczone dla miar zanurzania sympleksowego Liu pozwoliły na utworzenie trzech klas województw ze względu na ich wartości. W tabeli 3 podano numery klas ( $KL_{2005}$ ,  $KL_{2010}$ ,  $KL_{2015}$ ), do których należą dane województwa. Można przyjąć, że jednostki należące do tych samych klas są podobnie oddalone od centralnego skupienia zbiorów danych, wyznaczonego przez mediany zanurzania Liu.

Województwa podkarpackie i warmińsko-mazurskie we wszystkich badanych latach należą do klasy pierwszej. Są one najbardziej oddalone od centralnego skupienia zbiorów danych ze względu na najmniejszą liczbę przestępstw przeciwko mieniu przypadającą na 1000 mieszkańców (województwo podkarpackie) oraz najwyższe wartości stopy bezrobocia (województwo warmińsko-mazurskie). Województwom dolnośląskiemu i lubelskiemu, które w 2005 roku należały do drugiej klasy, w kolejnych latach odpowiadały najniższe wartości miary zanurzania. Charakteryzują się one znacznym spadkiem stopy bezrobocia (województwo dolnośląskie) oraz małą liczbą przestępstw przeciwko mieniu (województwo lubelskie). Najwyższe wartości miary zanurzania we wszystkich badanych latach odpowiadają województwom łódzkiemu i opolskiemu. Położone są one najbardziej centralnie w „chmurze danych”. Można przyjąć, że wartości liczbowe wskaźnika przestępstw przeciwko mieniu oraz stopy bezrobocia osiągnęły w tych województwach w danych latach najbardziej typowe wartości.

Zanurzanie Tukeya zostało wykorzystane do sporządzenia wykresów konturów zanurzania (rysunek 1) oraz wykresów pudełkowych sporządzonych dla danych dwuwymiarowych (rysunek 2).

Wykresy te pozwalają na uchwycenie kilku charakterystycznych własności zbiorów danych. Kształt i orientacja konturów zanurzania wskazują na ujemną korelację pomiędzy zmiennymi w 2010 i 2015 roku, przy czym silniejsza jest ona w ostatnim badanym roku. Kształt konturów zanurzania w 2005 roku jest bardziej okrągły w porównaniu z pozostałymi latami, korelacja pomiędzy zmiennymi w tym przypadku jest najmniejsza. Wyznaczone współczynniki korelacji liniowej Pearsona potwierdzają fakt, że najsilniejsza zależność liniowa ujemna pomiędzy stopą bezrobocia i liczbą przestępstw przeciwko mieniu występowała w 2015 roku ( $r = -0,4796$ ), w poprzednim roku współczynnik ten wynosił  $r = -0,1896$ .

Zauważyć można, że mediana umieszczona jest centralnie na wykresach pudełkowych dwuwymiarowych. Kontury zanurzania pozwalają na dokonanie

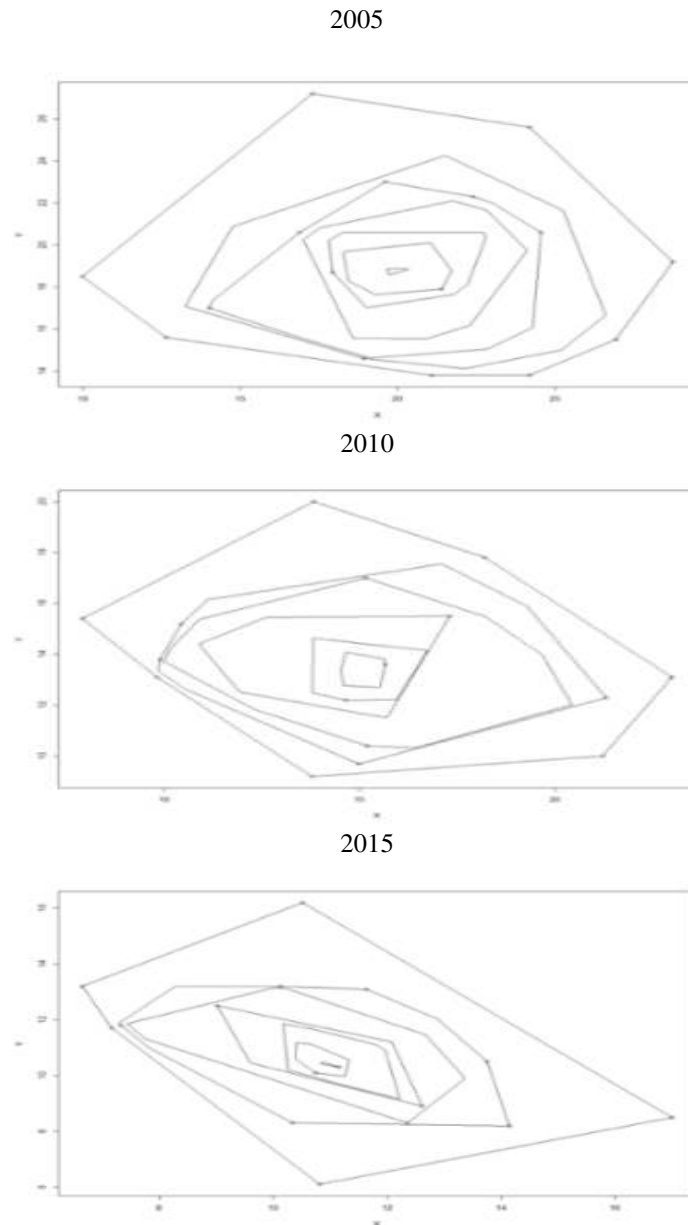
Tabela 3. Miary zanurzenia Tukeya oraz simpleksowego Liu

<i>W</i>	$T_{zan,2003}$	$RT_{2003}$	<i>W</i>	$L_{zan,2003}$	$KL_{2003}$	<i>W</i>	$T_{zan,2010}$	$RT_{2010}$	<i>W</i>	$L_{zan,2010}$	$KL_{2010}$	<i>W</i>	$T_{zan,2013}$	$RT_{2013}$	<i>W</i>	$L_{zan,2013}$	$KL_{2013}$
6	0,000	4,5	6	0,188	1	1	0,000	4	1	0,188	1	1	0,000	3	1	0,188	1
7	0,000	4,5	7	0,188	1	3	0,000	4	3	0,188	1	3	0,000	3	3	0,188	1
9	0,000	4,5	9	0,188	1	9	0,000	4	9	0,188	1	9	0,000	3	9	0,188	1
10	0,000	4,5	10	0,188	1	12	0,000	4	12	0,188	1	14	0,000	3	14	0,188	1
11	0,000	4,5	11	0,188	1	14	0,000	4	14	0,188	1	15	0,000	3	15	0,188	1
12	0,000	4,5	12	0,188	1	15	0,000	4	15	0,188	1	2	0,063	8,5	12	0,211	2
14	0,000	4,5	14	0,188	1	16	0,000	4	16	0,188	1	4	0,063	8,5	10	0,230	2
16	0,000	4,5	16	0,188	1	7	0,063	9,5	11	0,211	2	6	0,063	8,5	4	0,246	2
15	0,063	9	15	0,211	2	10	0,063	9,5	7	0,230	2	10	0,063	8,5	16	0,246	2
1	0,125	12	1	0,279	2	11	0,063	9,5	10	0,246	2	12	0,063	8,5	7	0,250	2
2	0,125	12	4	0,288	2	13	0,063	9,5	13	0,259	2	16	0,063	8,5	2	0,259	2
3	0,125	12	3	0,295	2	2	0,125	12,5	2	0,288	2	7	0,125	12	6	0,259	2
4	0,125	12	13	0,295	3	6	0,125	12,5	4	0,298	3	11	0,188	13,5	13	0,311	3
13	0,125	12	2	0,311	3	4	0,188	14	6	0,316	3	13	0,188	13,5	11	0,323	3
8	0,250	15	8	0,395	3	5	0,250	15	5	0,386	3	8	0,313	15	8	0,411	3
5	0,313	16	5	0,413	3	8	0,313	16	8	0,425	3	5	0,375	16	5	0,430	3

Źródło: opracowanie własne na podstawie obliczeń w pakietach „depthproc” i „mrfdepth

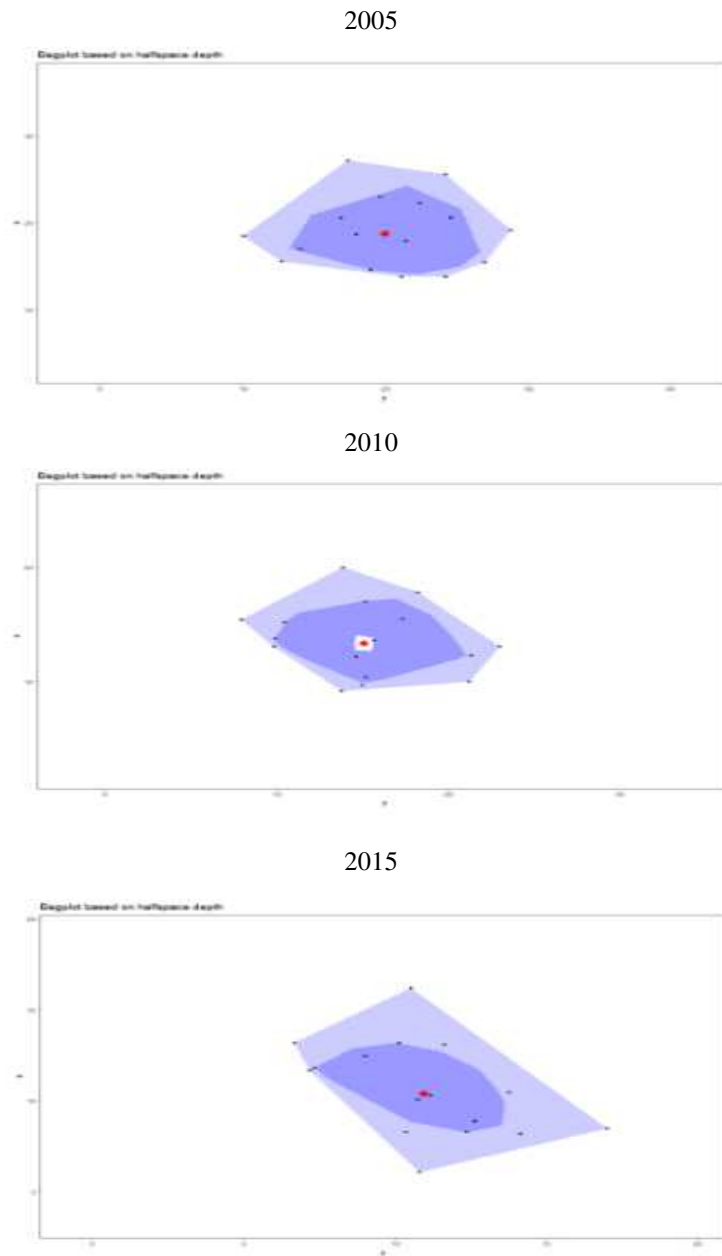
rangowania województw względem oddalenia od centralnego skupienia zbiorów danych. W tabeli 3 zamieszczono rangi województw ze względu na odpowiadającą im wartość miary zanurzania Tukeya ( $RT_{2005}$ ,  $RT_{2010}$ ,  $RT_{2015}$ ).

Rysunek 1. Kontury zanurzania Tukeya



Źródło: opracowanie własne

Rysunek 2. Wykresy pudełkowe dla danych dwuwymiarowych



Źródło: opracowanie własne



## PODSUMOWANIE

Na podstawie przeprowadzonych analiz można stwierdzić, że zauważalne jest znaczne różnicowanie poziomu badanych cech w województwach, przy czym największe odnotowano w 2005 roku. Niższy wskaźnik zagrożenia przestępczością przeciwko mieniu przypadający na 1000 mieszkańców występuje w województwach wschodnich Polski. W porównaniu z województwami dolnośląskim, lubuskim, śląskim i zachodniopomorskim wskaźnik ten był ponad dwukrotnie niższy. Duże aglomeracje miejskie oraz województwa zlokalizowane w bliskim sąsiedztwie zachodniej granicy wyróżniają się natężeniem przestępstw powyżej przeciętnej w kraju<sup>4</sup>.

Bezrobociu przypisuje się ważną rolę w kształtowaniu poziomu przestępczości. Może przybierać różne kierunki obniżając lub podnosząc jego poziom. Na wskaźnik przestępczości oprócz bezrobocia mogą mieć wpływ inne czynniki o charakterze społeczno-ekonomicznym [Szczepaniec 2012].

W pracy zaprezentowano wykorzystanie koncepcji zanurzania obserwacji w próbie do analizy danych liczbowych dotyczących liczby przestępstw przeciwko mieniu oraz stopy bezrobocia. Dokonano rangowania jednostek względem oddalenia od centralnego skupienia zbiorów danych, co pozwoliło na wyodrębnienie województw, które można uznać za typowe ze względu na wartości cech diagnostycznych oraz tych, w których cechy osiągnęły w badanych latach znacznie wyższe lub niższe wartości. Techniki graficzne oparte na zanurzeniu Tukeya obserwacji w próbie umożliwiły zobrazowanie pewnych własności zbiorów danych oraz uchwycenie ich centralnego skupienia.

W pracy zaprezentowano przykłady zastosowania miar i konturów zanurzania w analizie danych dwuwymiarowych. Zagadnienia te będą przedmiotem dalszych badań.

## BIBLIOGRAFIA

- Chebana F., Ouarda T. B. M. J. (2011) Depth-Based Multivariate Descriptive Statistics with Hydrological Applications. *Journal of Geophysical Research*, 116, D1020, doi:10.1029/2010JD015338.
- Donoho D. L., Gasko M. (1992) Breakdown Properties of Location Estimates Based on Halfspace Depth and Projected Outlyingness. *The Annals of Statistics*, 20, 1803-1827.
- Hołyst B. (2016) *Kryminologia*. Wolters Kluwer Polska SA, Warszawa.
- Kędziołka K. (2014) Wpływ wybranych czynników o charakterze społeczno-ekonomicznym na przestępczość przeciwko mieniu w Polsce. *Stud. Ekonom.* 181, 11-23.

---

<sup>4</sup> <http://wiadomosci.onet.pl/kraj/rzeczpospolita-wzrasta-przestepczosc-najgorzej-na-zachodzie-polski/rc7rj> [dostęp: 1.08.2017]

- Kobylińska M. (2006) Comparison of Selected Criteria for Determination of the Measure of Depth of an Observation in a Two-Dimensional Sample. *Acta Universitatis Lodzianis, Folia Oeconomica*, 196, 139-153.
- Kosiorowski D. (2012) *Statystyczne funkcje głębi w odpornej analizie ekonomicznej*. Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie.
- Liu R. Y. (1990) On a Notion of Data Depth Based on Random Simplices. *The Annals of Statistics*, 18, 405-414.
- Liu R. Y., Parelius J. M., Singh K. (1999) Multivariate Analysis by Data Depth: Descriptive Statistics, Graphics and Inference. *The Annals of Statistics*, 27, 783-858.
- Mosler K. (2002) *Multivariate Dispersion, Central Regions and Depth: the Lift Zonoid Approach*. Springer, New York.
- Rousseeuw P. J., Ruts I., Tukey J. W. (1999) The Bagplot: A Bivariate Boxplot. *The American Statistician*, 53 (4), 382-387.
- Rousseeuw P. J., Ruts I. (1996) Bivariate Location Depth. *Applied Statistics*, 45, 516-526.
- Szczepaniec M. (2011) Kontrowersje wokół wpływu warunków ekonomicznych na poziom przestępczości. *Czasopismo Prawa Karnego i Nauk Penalnych*, 15(4), 107-123.
- Szczepaniec M. (2012) Bezrobocie jako czynnik kształtujący czynnik przestępczości. *Zeszyty Prawnicze*, 12(3), 165-176.
- Sztaudynger J. J., Sztaudynger M. (2003) Ekonometryczne modele przestępczości. *Ruch Prawniczy, Ekonomiczny i Socjologiczny*, 65(3), 127-143.
- Tukey J. (1975) Mathematics and the Picturing of Data. *Proceedings of the International Congress of Mathematicians, Vancouver*, 2, 523-531.
- Wagner W., Kobylińska M. (2002) Przegląd metod wyznaczania miar i konturów zanurzenia w próbie dwuwymiarowej. *Przebieg Statystyczny*, 49(4), 119-131.
- Zawłocki R. (red.) (2011) *System prawa karnego*. Wydawnictwo C. H. Beck, Warszawa, 9.
- Zuo Y., Serfling R. (2000) General Notations of Statistical Depth Function. *Annals Statistics*, 28, 461-482.
- <https://bdl.stat.gov.pl/BDL/dane/podgrup/temat> [dostęp: 10.06.2017] (dane dotyczące stopy bezrobocia oraz liczby przestępstw przeciwko mieniu).
- <https://cran.r-project.org/web/packages/depth/index.html> [dostęp: 15.06.2017].
- <https://cran.r-project.org/web/packages/DepthProc/index.html> [dostęp: 15.06.2017].
- <https://cran.r-project.org/web/packages/mrfDepth/index.html> [dostęp: 15.06.2017].
- <http://wiadomosci.onet.pl/kraj/rzeczpospolita-wzrasta-przestepczosc-najgorzej-na-zachodzie-polski/rc7rj> [dostęp: 1.08.2017].

**OBSERVATION DEPTH MEASURE IN A SAMPLE, APPLIED  
TO ASSESSMENT OF THE VARIETY OF CRIMES AGAINST  
PROPERTY AND THE UNEMPLOYMENT RATE IN POLAND**

**Abstract:** This article presents the application of an observation depth measure in a sample to analyse data related to the number of crimes committed against property and the unemployment rate. The provinces were ranked based on their distance from the central data cluster. Graphs based on observation depth measures in a sample allowed for visualisation of certain properties of these data.

**Keywords:** observation depth measure in a sample, Tukey's depth measure, Liu's simplex depth measure, depth contours

## WPLYW ZMIAN FRAKCJI KOBIET W MANAGEMENTCIE NA ZMIANY ROE SPÓŁEK PUBLICZNYCH<sup>1</sup>

**Krzysztof Kompa (ORCID: 0000-0002-2810-6654)**

Wydział Zastosowań Informatyki i Matematyki  
Szkola Główna Gospodarstwa Wiejskiego w Warszawie  
e-mail: krzysztof\_kompa@sggw.pl

**Dorota Witkowska (ORCID: 0000-0001-9538-9589)**

Wydział Zarządzania  
Uniwersytet Łódzki  
e-mail: dorota.witkowska@uni.lodz.pl

**Streszczenie:** Parytet 30%-40% udziału kobiet w organach kierowniczych dużych spółek publicznych został wprowadzony w wielu krajach UE. W uzasadnieniu wskazuje się na poprawę wyników finansowych firm z kobietami w managementcie. Celem pracy jest ocena sytuacji w Polsce, badanie zmian frakcji kobiet w kierownictwach spółek notowanych na GPW i analiza wpływu tych zmian na sytuację finansową spółek. Badaniem objęto 32 spółki z indeksu WIG30 i spółki Skarbu Państwa. Otrzymane wyniki nie potwierdzają dodatniej korelacji między zmianami frakcji kobiet, a zmianami ROE.

**Słowa kluczowe:** spółki publiczne, rentowność kapitału własnego, obecność kobiet w kierownictwie

### WSTĘP

Dyskusje na temat potrzeby uwzględniania parytetu płci w kierownictwie przedsiębiorstw są prowadzone od ponad dekady. Przytacza się przy tym różne argumenty wspierające tę ideę. Są to zarówno argumenty równościowe tj. niedopuszczalność dyskryminacji członków organów statutowych ze względu na płeć, jak i ekonomiczne, wg których sytuacja finansowa spółki poprawia się wraz ze wzrostem reprezentacji kobiet w organach zarządczych (np.: [Catalyst 2004,

---

<sup>1</sup> Badania realizowano w ramach projektu badawczego UMO-2015/17/B/HS4/00930 sfinansowanego ze środków Narodowego Centrum Nauki.

Campbell, Minguez-Vera 2008, McKinsley 2007, Lisowska i in. 2014]). Nie brak też opracowań przeczących temu twierdzeniu (por. prace: [Lee, James 2007, Ahern, Dittmar 2012]) lub wskazujących na brak związku (np. [Wang, Clift 2009 i Kompa i in. 2016]).

Celem badania jest ocena zmian udziału kobiet w organach kierowniczych spółek publicznych w Polsce oraz sprawdzenie występowania zależności między zmianami rentowności kapitału własnego (ROE) i zmianami reprezentacji kobiet w zarządach i radach nadzorczych spółek Skarbu Państwa oraz należących do indeksu giełdowego WIG30, w latach 2010-2015.

## KOBIETY W KIEROWNICTWIE SPÓŁEK PUBLICZNYCH W POLSCE

Warto przypomnieć, że w 2012 r. Komisja Europejska przedstawiła wniosek w sprawie dyrektywy o 40% reprezentacji kobiet w radach nadzorczych spółek publicznych i skarbu państwa, które spełniają kryteria zaliczające je do dużych firm. W grudniu 2015 r. projekt dyrektywy został odrzucony przez ministrów państw UE wobec argumentacji, że wiele krajów posiada własne regulacje (lub jest w trakcie ich tworzenia), poprawiające równowagę płci w organach spółek. Przyjrzyjmy się zatem na ile sytuacja w Polsce odbiega od proponowanej struktury.

Tabela 1. Udział kobiet (%) w organach kierowniczych spółek notowanych na GPW

Rok	Zarząd ogółem			Zarząd wg funkcji					Rada Nadzorcza wg funkcji				
	GPW	RG	NC	(1)		(2)	(3)		(1)			ogółem	(1)
				RG	NC		GPW			GPW		RG	
2010	11,1	10,8	11,0	6,3	9,6	7,1	10,0	16,1	11,7	9,1	21,3	10,7	9,1
2011	11,9	10,6	14,2	6,8	10,7	7,9	11,2	16,8	10,2	8,2	15,3	10,6	8,2
2012	13,0	12,0	14,1	6,5	8,3	7,3	13,9	18,7	12,8	8,5	18,5	11,2	8,5
2013	12,5	11,5	14,5	6,7	9,2	7,9	12,4	18,1	12,5	8,9	16,7	11,4	8,9
2014	12,7	12,0	14,0	7,5	6,8	7,2	12,4	19,8	12,8	9,1	16,9	13,3	9,1
2015	12,4	11,6	14,0	7,6	7,3	7,4	12,3	18,5	13,3	9,9	17,1	14,2	9,9
2016	12,0	11,1	13,8	6,9	7,2	7,0	11,6	18,3	11,7	7,8	17,1	15,1	7,8

Uwaga: cyfry (1), (2), (3) oznaczają funkcje, odpowiednio: prezesa zarządu lub przewodniczącego rady nadzorczej, zastępców i członków; GPW oznacza wszystkie rynki, RG – rynek główny, NC – rynek NewConnect.

Źródło: opracowanie własne na podstawie Notoria Serwis

Liczba zatrudnionych kobiet w spółkach publicznych jest znaczna, ale ich reprezentacja w managementcie jest dość uboga. W radach nadzorczych zazwyczaj zasiada więcej kobiet niż w zarządach, częściej też zajmują kierownicze stanowiska w organach doradczych. Biorąc pod uwagę procentowy udział kobiet w organach kierowniczych wszystkich spółek notowanych na GPW w latach 2010-2016 (tzn. zarówno na rynku głównym, jak i na NewConnect), zauważa się (tabela 1) nieznaczny: (a) wzrost ich procentowego udziału w zarządach - o 14% wzrosła liczba członków, o 16% liczba wiceprezesów, chociaż o 1% spadła liczba kobiet na

najwyższym stanowisku, (b) spadek udziału w radach nadzorczych: 20% spadek liczby członków i 14% wśród zastępców przewodniczącego. Występuje przy tym znaczne różnicowanie struktury organów kierowniczych w zależności od badanego rynku. Na rynku alternatywnym GPW (NewConnect – NC) obserwuje się 25,5% wzrost liczby kobiet w zarządach, przy jednoczesnym 25,4% spadku udziału kobiet wśród prezesów. Rynek główny GPW (RG) charakteryzuje sytuacja odmienna: odnotowuje się jedynie 9,5% wzrost prezesur i 2,5% wzrost frakcji kobiet w zarządach. Na całym rynku odnotowano 7,7% wzrost frakcji kobiet w zarządach, przy 1,2% spadku ich reprezentacji na stanowiskach prezesów.

Należy zauważyć, że w ciągu analizowanych siedmiu lat udział kobiet w zarządach (tabela 1) zwiększył się jedynie o niecałe 3% (w porównaniu do roku 2012 odnotowuje się spadek o 7,5%). W tym samym czasie frakcja w radach nadzorczych wzrosła o ponad 41% (35% w relacji do roku 2012), co daje jednak tylko 15% udział kobiet i zasadniczo odbiega od zaleceń Komisji Europejskiej.

Tabela 2. Procentowy udział kobiet w organach kierowniczych spółek notowanych na głównym parkiecie

Rady nadzorcze			Rady nadzorcze ogółem							
			wg indeksów			wg sektorów				
Rok	ogółem	Przewodniczący	Rok	WIG20	mWIG40	sWIG80	Przemysł	Finanse	Usługi	
2010	10,7	9,1	2014	17,6	10,3	11,3	16,5	14,8	18,9	
2011	10,6	8,2	2015	19,9	12,3	11,8	18,0	15,7	18,7	
2012	11,2	8,5	2016	18,3	11,2	14,2	18,5	15,3	20,5	
2013	11,4	8,9	Zarządy ogółem							
2014	13,3	9,1	2014	5,3	12,1	11,5	10,5	15,1	13,0	
2015	14,2	9,9	2015	11,8	7,2	11,9	9,5	15,1	13,2	
2016	15,1	7,8	2016	12,6	6,7	11,3	9,3	14,4	12,4	

Źródło: opracowanie własne na podstawie Notoria Serwis

Analiza frakcji kobiet w spółkach o różnej kapitalizacji (reprezentowanych przez trzy indeksy giełdowe) i z trzech sektorów wyróżnianych przez GPW wskazuje, że zachodzi nieznaczny wzrost udziału kobiet w radach nadzorczych. Widoczny jest też spadek frakcji w zarządach spółek notowanych w mWIG40, stabilizacja w spółkach z sWIG80 oraz 140% wzrost w spółkach największych (tabela 2).

## POMIAR ANALIZOWANYCH ZJAWISK

Podstawowymi kwestiami, jakie należy rozwiązać przy badaniu wpływu zróżnicowania struktury managementu na efektywność finansową przedsiębiorstw, jest pomiar obu czynników. Literatura przedmiotu nie poświęca wiele miejsca tym zagadnieniom. W przypadku oceny sytuacji finansowej przedsiębiorstw najczęściej korzysta się z takich wskaźników jak np.: ROE, ROA, całkowita stopa zwrotu dla

akcjonariuszy, zwrot z zainwestowanego kapitału, wartość rynkowa, a także wykorzystuje się z mierniki syntetyczne (por. [Kompa i in. 2016] czy [Kmieciak 2017]). Udział kobiet w strukturach kierowniczych określa się zazwyczaj procentowo, chociaż stosuje się również pewne wskaźniki heterogeniczności, np. tzw. indeks Blaua [Blau 1997]. Natomiast rzadko uwzględnia się aspekt czasu, związany z opóźnieniem skutków decyzji zarządczych. Innymi słowy należy uwzględnić opóźnienie pomiędzy zmianą struktury managementu, a pomiarem wyniku finansowego, bo efekt gospodarczy zmiany sposobu zarządzania spółką nie jest natychmiastowy. Zarazem obserwacje tworzą szeregi momentów, co skutkuje koniecznością dobru odpowiednich punktów pomiarowych<sup>2</sup>.

Na potrzeby tego badania przyjęto, że dane o składach zarządów i rad nadzorczych brane są na dzień 30. czerwca każdego roku, a rentowność kapitału własnego (ROE) mierzona jest na koniec roku. Badano spółki z kompozycji indeksu WIG30 w listopadzie 2016 roku i spółki Skarbu Państwa na podstawie danych Notoria Serwis. Ostatecznie uwzględniono 32 spółki – 9 finansowych i 23 niefinansowe. Pełna lista wraz z oznaczeniami została zamieszczona w tabeli w załączniku.

## STRUKTURA ZATRUDNIENIA W ORGANACH KIEROWNICZYCH

Specyfikę zatrudnienia opisano wskaźnikami struktury i ich przyrostami. Biorąc pod uwagę wszystkie analizowane spółki w badanym okresie sześcioletnim (tabela 3) zauważa się 39% (o 3,13 p.p.) wzrost udziału kobiet w organach kierowniczych. Dla spółek finansowych przyrost względny frakcji jest o ponad 10 p.p. większy niż w drugiej grupie przedsiębiorstw. Zwraca też uwagę (tabele 4 i 5) występowanie przedsiębiorstw (w każdym roku analizy i w każdej grupie spółek) bez kobiet w organach kierowniczych. W spółkach niefinansowych maksymalny udział pań wynosi 40% i ok. 36%, odpowiednio dla spółek Kernel (N10), rok 2010 i Tauron (N21), rok 2015. Natomiast w instytucjach finansowych największa frakcja kobiet (40%) wystąpiła w roku 2015 (Alior Bank, F2) i w roku 2010, gdy najwięcej kobiet było w managementcie Giełdy Papierów Wartościowych (F1) – 36%.

Tabela 3. Średni procentowy udział kobiet w zarządach i radach nadzorczych badanych spółek w latach 2010-2015

Spółki	Udział procentowy w latach						Przyrosty w latach 2015-2010	
	2010	2011	2012	2013	2014	2015	bezwzględne	względne
niefinansowe	9,35	10,04	10,99	10,88	10,00	12,50	3,15	33,65%
finansowe	12,21	11,59	11,11	13,97	16,43	17,81	5,60	45,80%
wszystkie	10,27	10,55	11,03	11,88	12,09	14,29	4,02	39,12%

Źródło: opracowanie własne na podstawie Notoria Serwis

<sup>2</sup> Wspomniane aspekty uwzględniono w pracach [Kompa i in. 2016, Kmieciak 2017].

Tabela 4. Frakcje kobiet w kierownictwach spółek niefinansowych w latach 2010-2015

Spółki	Procentowe udziały w latach						Przyrosty bezwzględne w latach					
	2010	2011	2012	2013	2014	2015	2011	2012	2013	2014	2015	2015/ 2010
N1	5,00	5,56	4,76	5,00	0,00	0,00	0,56	-0,79	0,24	-5,00	0,00	-5,00
N2	25,00	16,67	15,38	7,69	6,67	6,67	-8,33	-1,28	-7,69	-1,03	0,00	-18,30
N3	10,00	10,00	9,09	9,09	11,11	0,00	0,00	-0,91	0,00	2,02	-11,11	-10,00
N4	14,29	10,00	10,00	10,00	10,00	9,09	-4,29	0,00	0,00	0,00	-0,91	-5,19
N5	11,11	11,11	11,11	11,11	9,09	9,09	0,00	0,00	0,00	-2,02	0,00	-2,02
N6	7,14	16,67	15,38	23,08	25,00	27,27	9,52	-1,28	7,69	1,92	2,27	20,13
N7	8,33	8,33	8,33	8,33	8,33	7,69	0,00	0,00	0,00	0,00	-0,64	-0,64
N8	0,00	0,00	0,00	5,56	5,56	5,26	0,00	0,00	5,56	0,00	-0,29	5,26
N9	0,00	0,00	0,00	0,00	0,00	12,50	0,00	0,00	0,00	0,00	12,50	12,50
N10	40,00	28,57	25,00	25,00	25,00	25,00	-11,43	-3,57	0,00	0,00	0,00	-15,00
N11	0,00	6,25	15,38	16,67	7,14	7,69	6,25	9,13	1,28	-9,52	0,55	7,69
N12	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
N13	16,67	10,00	22,22	18,18	27,27	25,00	-6,67	12,22	-4,04	9,09	-2,27	8,33
N14	10,00	10,00	9,09	0,00	0,00	11,11	0,00	-0,91	-9,09	0,00	11,11	1,11
N15	5,56	11,11	5,56	9,52	15,00	21,74	5,56	-5,56	3,97	5,48	6,74	16,18
N16	14,29	16,67	25,00	33,33	30,77	23,08	2,38	8,33	8,33	-2,56	-7,69	8,79
N17	16,67	16,67	30,77	25,00	18,18	9,09	0,00	14,10	-5,77	-6,82	-9,09	-7,58
N18	6,67	13,33	6,67	8,33	7,14	7,14	6,67	-6,67	1,67	-1,19	0,00	0,48
N19	13,33	9,09	8,33	13,33	0,00	7,14	-4,24	-0,76	5,00	-13,33	7,14	-6,19
N20	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
N21	15,38	15,38	15,38	16,67	14,29	35,71	0,00	0,00	1,28	-2,38	21,43	20,33
N22	11,11	10,00	11,11	11,11	20,00	30,00	-1,11	1,11	0,00	8,89	10,00	18,89
N23	11,11	22,22	22,22	11,11	0,00	0,00	11,11	0,00	-11,11	-11,11	0,00	-11,10
Min.	0,00	0,00	0,00	0,00	0,00	0,00	-11,43	-6,67	-11,11	-13,33	-11,11	-18,30
Max.	40,00	28,57	30,77	33,33	30,77	35,71	11,11	14,10	8,33	9,09	21,43	20,33

Źródło: opracowanie własne na podstawie Notoria Serwis

Tabela 5. Frakcje kobiet w kierownictwach spółek finansowych w latach 2010-2015

Spółki	Procentowe udziały w latach						Przyrosty bezwzględne w latach					
	2010	2011	2012	2013	2014	2015	2011	2012	2013	2014	2015	2015/ 2010
F1	36,36	18,18	18,18	8,33	0,00	0,00	-18,18	0,00	-9,85	-8,33	0,00	-36,36
F2	20,00	25,00	25,00	27,27	40,00	40,00	5,00	0,00	2,27	12,73	0,00	20,00
F3	5,00	4,35	0,00	0,00	10,00	12,50	-0,65	-4,35	0,00	10,00	2,50	7,50
F4	21,43	25,00	23,08	23,08	21,43	20,00	3,57	-1,92	0,00	-1,65	-1,43	-1,43
F5	11,76	11,76	5,88	11,76	15,79	15,79	0,00	-5,88	5,88	4,02	0,00	4,02
F6	5,26	10,53	10,53	15,79	15,79	16,67	5,26	0,00	5,26	0,00	0,88	11,40
F7	6,25	6,67	21,43	26,67	26,67	25,00	0,42	14,76	5,24	0,00	-1,67	18,75
F8	0,00	6,25	6,67	13,33	12,50	20,00	6,25	0,42	6,67	-0,83	7,50	20,00
F9	20,00	7,69	0,00	6,25	13,33	7,69	-12,31	-7,69	6,25	7,08	-5,64	-12,31
Min.	0,00	0,00	0,00	0,00	0,00	0,00	-18,18	-7,69	-9,85	-8,33	-5,64	-36,36
Max.	36,36	25,00	25,00	27,27	36,36	40,00	6,25	14,76	6,67	12,73	7,50	20,00

Źródło: opracowanie własne na podstawie Notoria Serwis



## OCENA SYTUACJI FINANSOWEJ SPÓŁEK

Do oceny sytuacji finansowej badanych spółek wykorzystywano wskaźnik rentowności kapitału własnego (ROE), który jest często stosowany w jednowymiarowych analizach porównawczych przedsiębiorstw. Dane dla obu analizowanych grup spółek wraz z dynamiką zmian zestawiono w tabelach 6 i 7. Wskaźniki rentowności są dodatnie dla niemal wszystkich analizowanych spółek we wszystkich latach badania. Wyjątkiem są spółki: Globe Trade Centre (N8) - za lata 2011-2014, CD Projekt (N4) i Allior Bank (F2) za 2010 rok oraz ENEA (N6) za 2015 rok. Nie ma natomiast spółki, która nie odnotowałaby spadków współczynnika ROE z okresu na okres (choćby jednokrotnego). Największe zróżnicowanie wartości ROE widoczne jest w przypadku: (a) spółek niefinansowych za rok 2015 (od -58,66 do 100,46) oraz (b) spółek finansowych w 2010 roku (od -10,13 do 20,27). Podobne obserwacje dotyczą przyrostów bezwzględnych wskaźnika ROE.

Tabela 6. Spółki niefinansowe: ROE, przyrosty bezwzględne i dynamika za lata 2010-2015

Spółki	ROE						Przyrosty bezwzględne w latach					
	2010	2011	2012	2013	2014	2015	2011	2012	2013	2014	2015	2015/ 2010
N1	9,29	9,07	7,81	6,14	6,86	7,06	-0,22	-1,26	-1,67	0,72	0,19	-2,23
N2	26,03	16,78	9,57	12,17	4,02	9,57	-9,25	-7,21	2,60	-8,15	5,56	-16,46
N3	30,89	26,75	20,77	22,35	54,45	24,99	-4,14	-5,98	1,58	32,10	-29,46	-5,90
N4	-8,49	23,44	20,46	9,09	2,79	100,46	31,93	-2,98	-11,37	-6,29	97,67	108,96
N5	68,89	13,79	27,42	19,22	4,84	12,04	-55,11	13,63	-8,20	-14,37	7,19	-56,86
N6	6,64	7,80	6,49	6,45	7,72	-3,30	1,16	-1,31	-0,04	1,27	-11,02	-9,94
N7	31,18	26,73	37,77	26,60	19,04	20,86	-4,45	11,04	-11,17	-7,56	1,82	-10,32
N8	2,73	-39,84	-17,40	-27,55	-41,65	8,24	-42,56	22,44	-10,15	-14,11	49,90	5,52
N9	27,55	28,68	11,61	0,97	-8,41	-58,66	1,13	-17,07	-10,64	-9,39	-50,24	-86,21
N10	31,53	28,22	19,09	8,21	-9,01	9,94	-3,31	-9,12	-10,88	-17,23	18,96	-21,58
N11	37,11	58,09	20,98	13,49	10,09	-21,80	20,98	-37,11	-7,50	-3,40	-31,89	-58,91
N12	19,36	32,76	33,39	31,97	30,74	19,92	13,40	0,63	-1,42	-1,23	-10,83	0,55
N13	9,49	8,49	11,01	0,43	-16,81	-3,30	-1,00	2,52	-10,58	-17,24	13,51	-12,79
N14	12,41	10,76	13,92	13,88	10,94	-12,04	-1,65	3,16	-0,04	-2,94	-22,98	-24,45
N15	0,69	13,24	6,27	2,30	4,28	2,08	12,55	-6,98	-3,97	1,98	-2,19	1,39
N16	9,42	12,53	8,87	9,36	8,25	-7,12	3,11	-3,66	0,49	-1,12	-15,37	-16,54
N17	10,93	7,20	8,55	6,90	9,63	7,01	-3,73	1,35	-1,65	2,73	-2,61	-3,92
N18	10,69	7,90	7,88	0,32	-24,32	14,49	-2,79	-0,02	-7,55	-24,64	38,80	3,80
N19	2,59	15,16	8,62	1,95	2,27	0,93	12,57	-6,54	-6,68	0,32	-1,34	-1,66
N20	25,20	37,90	19,95	15,99	15,77	18,76	12,70	-17,95	-3,96	-0,22	2,99	-6,44
N21	6,73	8,10	9,43	7,78	6,63	-10,60	1,36	1,34	-1,65	-1,16	-17,22	-17,33
N22	4,77	43,37	10,98	4,39	5,63	-6,83	38,60	-32,39	-6,59	1,24	-12,46	-11,60
N23	15,98	13,57	11,56	-10,77	15,34	11,23	-2,41	-2,01	-22,33	26,11	-4,11	-4,75
Min.	-8,49	-39,84	-17,40	-27,55	-41,65	-58,66	-55,11	-37,11	-22,33	-24,64	-50,24	-86,21
Max.	68,89	58,09	37,77	31,97	54,45	100,46	38,60	22,44	2,60	32,10	97,67	108,96

Źródło: opracowanie własne na podstawie Notoria Serwis

Tabela 7. Spółki finansowe: ROE, przyrosty bezwzględne i dynamika za lata 2010-2015

Spółki	ROE						Przyrosty bezwzględne w latach					
	2010	2011	2012	2013	2014	2015	2011	2012	2013	2014	2015	2015/ 2010
F1	18,12	25,57	19,67	19,02	16,78	17,40	7,45	-5,90	-0,65	-2,24	0,63	-0,72
F2	-10,13	14,58	3,97	10,97	12,39	9,46	24,71	-10,60	7,00	1,42	-2,92	19,60
F3	16,25	17,21	17,77	17,17	12,59	13,13	0,97	0,56	-0,60	-4,59	0,54	-3,12
F4	14,29	14,58	11,44	11,47	10,91	10,66	0,29	-3,15	0,03	-0,56	-0,24	-3,63
F5	12,00	15,00	14,00	12,00	12,00	11,00	3,00	-1,00	-2,00	0,00	-1,00	-1,00
F6	9,48	10,75	10,04	10,52	11,70	8,95	1,27	-0,72	0,48	1,18	-2,75	-0,53
F7	13,10	13,98	13,23	11,95	11,46	9,66	0,88	-0,75	-1,28	-0,49	-1,80	-3,44
F8	15,37	17,22	15,82	13,02	12,29	8,99	1,85	-1,41	-2,80	-0,73	-3,30	-6,39
F9	20,27	18,26	23,98	24,05	22,57	16,53	-2,01	5,72	0,07	-1,48	-6,05	-3,75
Min.	-10,13	10,75	3,97	10,52	10,91	8,95	-2,01	-10,60	-2,80	-4,59	-6,05	-6,39
Max.	20,27	25,57	23,98	24,05	22,57	17,40	24,71	5,72	7,00	1,42	0,63	19,60

Źródło: opracowanie własne na podstawie Notoria Serwis

### WSPÓLZALEŻNOŚĆ MIĘDZY ZMIANĄ STRUKTURY ORGANÓW KIEROWNICZYCH A RENTOWNOŚCIĄ KAPITAŁU WŁASNEGO

W analizach przyjęto, że wpływ obecności kobiet w managementie na sytuację finansową spółek badany będzie poprzez poszukiwanie zależności między zmianą struktury organów kierowniczych, a zmianą ROE. Uwzględniono przyrosty bezwzględne i względne, przy czym - z uwagi na występowanie przypadków braku kobiet w kierownictwie - zamiast klasycznego podejścia do obliczenia względnych miar dynamiki wykorzystano przyrosty bezwzględne, odniesione do wartości średniej danej zmiennej za badany okres. Dla dwóch spółek, w których nie było kobiet w organach kierowniczych w całym okresie, przyjęto zerowe wartości współczynników względnych. Występowanie współzależności między analizowanymi zmiennymi badano wykorzystując współczynnik korelacji Pearsona oraz test istotności tego współczynnika dla przyrostów łańcuchowych i jednopodstawowych, z uwzględnieniem bieżących i opóźnionych o rok relacji między zmiennymi.

Jak można zauważyć (tabela 8), w przypadku wykorzystania przyrostów bezwzględnych wszystkie badane zależności są statystycznie istotne ( $\alpha=0,05$ ), chociaż w większości przypadków wartości współczynnika Pearsona są ujemne. Dla przyrostów względnych badane relacje - w grupach spółek niefinansowych i wszystkich spółek poddanych analizie - stały się statystycznie istotne. Można zatem przypuszczać, że w badanym okresie wzrost udziału procentowego kobiet w managementie skutkował obniżeniem się rentowności kapitału własnego badanych spółek. Należy też stwierdzić, że analizowane zależności bieżące są silniejsze niż te przesunięte o jeden okres, tzn. kiedy zmiany w udziale kobiet mierzone są za lata 2011-2014, a dla ROE w okresie 2012-2015. W przypadku

spółek finansowych obserwujemy zjawisko przeciwne, tj. opóźnienie działa na nieznaczne zwiększenie wartości współczynnika korelacji.

Tabela 8. Wartości współczynników Pearsona ( $r$ ) oraz statystyki testowej ( $t$ ) wyznaczone dla przyrostów reprezentacji kobiet w kierownictwie i wskaźnika ROE

Spółki	Przyrosty łańcuchowe bezwzględne				Przyrosty jednopodstawowe bezwzględne	
	bieżące		opóźnione		bieżące	
	$r$	$t$	$r$	$t$	$r$	$t$
niefinansowe	-0,1164	-1,2453	-0,0181	-0,1719	-0,2113	-0,9907
niefinansowe*	-0,1101	-1,1246	0,0286	0,2589	-0,2179	-0,9730
finansowe	-0,0231	-0,1514	-0,1129	-0,7454	0,2238	1,5057
wszystkie	-0,1164	-1,4725	-0,0181	-0,2033	-0,2113	-1,1841
wszystkie*	-0,0958	-1,1705	-0,0182	-0,1980	-0,1129	-0,6011

Spółki	Przyrosty łańcuchowe względne				Przyrosty jednopodstawowe względne	
	bieżące		opóźnione		bieżące	
	$r$	$t$	$r$	$t$	$r$	$t$
niefinansowe	<b>-0,6714</b>	-9,6303	-0,0052	-0,0495	<b>-0,8097</b>	-6,3227
niefinansowe*	<b>-0,6712</b>	-9,1894	-0,0058	-0,0522	<b>-0,8089</b>	-5,9964
finansowe	-0,0320	-0,2098	-0,0684	-0,4496	0,0323	0,2119
wszystkie	<b>-0,6071</b>	-9,6030	0,0006	0,0070	<b>-0,6995</b>	-5,3612
wszystkie*	<b>-0,6069</b>	-9,2887	0,0007	0,0073	<b>-0,6987</b>	-5,1684

Objaśnienia: \* oznacza próby, z których wyłączono spółki N12 i N20; pogrubienie oznacza wartości statystycznie istotne na poziomie  $\alpha=0,05$ .

Źródło: obliczenia własne

## PODSUMOWANIE

Przedstawione analizy jednoznacznie potwierdzają, że odsetek kobiet w kierownictwie spółek publicznych w Polsce jest niski, mimo istotnego wzrostu reprezentacji kobiet w radach nadzorczych spółek z rynku głównego GPW. Zarazem efekty finansowe spółek zależą w głównej mierze od bieżących i - co najwyżej - niewiele opóźnionych decyzji zarządów, a w tych frakcja kobiet jest znikoma i bardzo odległa od tzw. masy krytycznej, określanej na poziomie 30%.

Uzyskane wyniki wskazują na to, że przyrost udziału kobiet zajmujących kierownicze pozycje w spółkach publicznych nie wpływa istotnie na wyniki finansowe mierzone wskaźnikiem ROE, lub jest istotnie ujemny, co stoi w sprzeczności z argumentami zwolenników wprowadzenia parytetu pci.

Otrzymane wyniki z pewnością obciąża celowy dobór spółek, wybrano bowiem tylko te, do których adresowana jest dyrektywa równościowa UE. Nie bez znaczenia wydaje się przy tym, że obsady organów kierowniczych, zwłaszcza spółek Skarbu Państwa, pochodzą z klucza politycznego, niekoniecznie uwzględniającego kompetencje mianowanych.

## BIBLIOGRAFIA

- Ahern K. R., Dittmar A. K. (2012) The Change of the Boards: The Impact on Firm Valuation of Mandated Female Board Representation. *The Quarterly Journal of Economics*, 127, 137-197, DOI: 101093/qje/qjr049.
- Blau P. M. (1997) *Inequality and Heterogeneity*. Free Press, New York, 1977.
- Campbell K., Minguez-Vera A. (2008) Gender Diversity in the Boardroom and Firm Financial Performance. *Journal of Business Ethics*, 83, 435-451.
- Catalyst (2004) *The Bottom Line: Connecting Corporate Performance and Gender Diversity*. Catalyst, New York.
- Kompa K., Mentel G., Witkowska D. (2016) Czy obecność kobiet we władzach spółek giełdowych wpływa na poprawę sytuacji finansowej tych spółek. *Metody Ilościowe w Badaniach Ekonomicznych*, XVII(3), 84-94.
- Kmieciak A. (2017) *Kobiety w zarządach spółek giełdowych*. Praca magisterska napisana pod kierunkiem D. Witkowskiej, Uniwersytet Łódzki, Łódź.
- Lee P. M., James E. H. J. (2007) She'-e-os: Gender Effects and Investor Reactions to the Announcements of Top Executive Appointments. *Strategic Management Journal*, 28(3), 227-241.
- Lisowska E., Zachorowska M., Sznajder A., Grabowska M. (2014) *Więcej kobiet w zarządzaniu – to się opłaca. Równość w biznesie, Raport na zlecenie MPiPS, Warszawa*.
- McKinsey (2007) *Women Matter: Gender Diversity, a Corporate Performance Driver*. McKinsey & Company, Inc.
- Wang Y., Clift B. (2009) Is There a Business Case for Board Diversity? *Pacific Accounting Review*, 21(2), 88-193.

## ZAŁĄCZNIK: LISTA SPÓŁEK I OZNACZENIA

Spółki niefinansowe - nazwa	Symbol	Spółki niefinansowe - nazwa	Symbol
Asseco Poland S.A.	N1	PKN Orlen S.A.	N18
Grupa Azoty S.A.	N2	PKP Cargo S.A.	N19
CCC S.A.	N3	Synthos S.A.	N20
CD Projekt S.A.	N4	Tauron Polska Energia S.A.	N21
Cyfrowy Polsat S.A.	N5	Zakłady Chemiczne Police S.A.	N22
Enea S.A.	N6	Elektrociepłownia Będzin S.A.	N23
Eurocash S.A.	N7		
Globe Trade Centre S.A.	N8	<b>Spółki finansowe - nazwa</b>	<b>Symbol</b>
Jastrzębska Spółka Węglowa S.A.	N9	GPW S.A.	F1
Kernel Holding S.A.	N10	Alior Bank S.A.	F2
KGHM Polska Miedź S.A.	N11	Bank Zachodni WBK S.A.	F3
LPP S.A.	N12	ING Bank Śląski S.A.	F4
Grupa Lotos S.A.	N13	Mbank S.A.	F5
Lubelski Węgiel Bogdanka S.A.	N14	Bank Millennium S.A.	F6
Orange Polska S.A.	N15	Bank Polska Kasa Opieki S.A.	F7
PGE S.A.	N16	PKO Bank Polski S.A.	F8
PGNiG S.A.	N17	PZU S.A.	F9

Źródło: opracowanie własne

**RELATION BETWEEN CHANGES OF WOMEN FRACTION  
IN MANAGEMENT AND ROE OF PUBLIC COMPANIES**

**Abstract:** The gender parity assuring 30-40% women in boards of big public companies, has been already affected in some EU states. Justification of this idea states the improvement of financial performance of companies with female managers. The aim of our research is evaluation of the situation in Poland, investigation of changes in women fraction in management of companies listed on WSE and analysis of its influence to changes of ROE. Research is provided for 32 companies from the stock index WIG30 or the State Treasury companies. The obtained results do not prove the positive correlation between changes of the women fraction and ROE.

**Keywords:** public companies, return on equity, women presence in boardrooms

## SKOŚNOŚĆ ROZKŁADU A ESTYMACJA KWANTYLOWYCH MIAR RYZYKA – PRZYPADEK RYNKU METALI

**Dominik Krężolek**

Wydział Informatyki i Komunikacji  
Uniwersytet Ekonomiczny w Katowicach  
e-mail: dominik.krezolek@ue.katowic.pl

**Streszczenie:** Pomiar ryzyka inwestycyjnego wymaga zastosowania narzędzi, które w odpowiedni sposób uwzględniają anomalie obserwowane w empirycznych rozkładach stóp zwrotu. Klasyczne modele szacowania ryzyka zakładają gaussowskie rozkłady prawdopodobieństwa, które nie uwzględniają asymetrii rozkładu, mającej związek z występowaniem obserwacji ekstremalnych. Takie obserwacje istotnie wpływają na poziom prawdopodobieństwa w ogonach rozkładów. W pracy podjęto próbę oceny wpływu skośności rozkładu prawdopodobieństwa na ocenę poziomu ryzyka inwestycji podejmowanych na rynku metali. Zastosowano kwantylowe miary ryzyka, m.in. wartość zagrożoną oraz warunkową wartość zagrożoną przy wykorzystaniu różnych teoretycznych rozkładów prawdopodobieństwa. Analizę przeprowadzono uwzględniając okres kryzysu finansowego.

**Słowa kluczowe:** skośność, pomiar ryzyka, Value-at-Risk, ryzyko ekstremalne, grube ogony

### WPROWADZENIE

Analityka rzeczywistości gospodarczej wymaga umiejętnego rozpoznania struktury badanego zjawiska. Niejednokrotnie można wskazać różne charakterystyki, które w istotnym stopniu mogą wpływać na wartości klasycznie wykorzystywanych miar deskrypcyjnych, a tym samym na interpretację wyników. Jednym z takich problemów jest zjawisko skośności obserwowane w rozkładzie wartości danych. Skośność w wartościach danych, a dokładniej asymetria, jest to sytuacja, kiedy występuje różnica pomiędzy podstawowymi miarami tendencji centralnej, tj. średnią, medianą i dominantą. Rozważając problem z punktu widzenia graficznej prezentacji danych, z asymetrią mamy do czynienia wtedy, gdy

kształt histogramu poniżej wartości centralnej jest różny od jego kształtu powyżej tej wartości.

W przypadku analizy zjawisk rzeczywistych nie występują rozkłady idealnie symetryczne, podobnie jak nie występują rozkłady idealnie normalne. Dlatego też ważnym zagadnieniem jest wskazanie, czy takie odchylenie od symetrii jest statystycznie istotne. W prezentowanym artykule podjęto próbę oceny wpływu skośności rozkładu na oszacowania wybranych kwantylowych miar ryzyka. Ta klasa miar jest niezwykle istotna z punktu widzenia wysokiego stopnia niepewności obserwowanej na rynkach, a co za tym idzie – istotnych zmian w poziomie cen (stóp zwrotu) analizowanych aktywów. Zjawiska zmienności i niepewności nie są przypisane do konkretnych rynków, są identyfikowalne na każdym z nich, oczywiście z odpowiednią intensywnością. W środowisku teoretyków oraz praktyków dominują analizy bazujące na informacjach płynących z rynków finansowych (przede wszystkim rynek kapitałowy). Teoretycy tworzą konstrukty analityczne w oparciu o dane rynkowe, natomiast praktycy wykorzystują te konstrukty zależnie od swoich potrzeb. Okazuje się, że wiele modeli klasycznie wykorzystywanych do opisu zjawisk na rynku akcji może być z powodzeniem wykorzystywanych również na innych rynkach, takich jak np. rynek towarowy.

Wszelkie modele oraz miary prezentowane w artykule zweryfikowano na danych rzeczywistych pochodzących z rynku metali. Eksploracja tego obszaru badawczego nie jest popularyzowana wśród analityków, a jeśli prowadzone są badania, to koncentrują się na analizie fundamentalnej, długookresowej. Brak jest prac badawczych dotyczących analizy ryzyka inwestycyjnego w metale, zatem prezentowany artykuł wypełnia po części tę lukę. Należy doprecyzować, iż przedmiotem analiz jest szczególna grupa metali, określana jako metale nieżelazne. Zaliczono do nich aluminium, miedź, ołów, nikiel, cynę oraz cynk. Są to metale stanowiące tzw. dodatki stopowe, uszlachetniające jakościowo produkt finalny a tym samym istotnie wpływające na jego cenę. Zastosowanie dodatku stopowego determinowane jest przeznaczeniem wyrobu gotowego. Metale wykorzystywane są w bardzo wielu dziedzinach działalności człowieka. Można tu wskazać m.in. wykorzystanie gospodarcze (przemysł konstrukcyjny, automotive, małe i duże AGD, przemysł wojskowy i kosmiczny, etc.) czy medyczne (stomatologia, chirurgia, kosmetyka, etc.). Dodatkowo metale notowane są na giełdach towarowych, zatem podlegają podobnym prawom i wpływom rynkowym jak inne aktywa (np. akcje). Stąd naturalnym zdaje się być wykorzystanie modeli popularyzowanych na innych, podobnych rynkach, także na rynku towarowym.

## METODOLOGIA

### Wybrane miary skośności danych

W literaturze przedmiotu problem skośności danych omawiany jest dość obszernie. Pierwsze wzmianki na ten temat pojawiają się na podstawowym kursie statystyki, a wraz ze wzrostem zaawansowania wiedzy wprowadzane są modyfikacje miar klasycznych celem wydobycia dodatkowej, często niejawniej informacji na temat źródła, kierunku czy siły asymetrii [Aczel 2000]. Dla zmiennej losowej  $X$  skośność rozkładu prawdopodobieństwa może być określona za pomocą trzeciego momentu centralnego jako  $M_3(X) = E([x - E(X)]^3)$ , wskazując tym samym kierunek asymetrii. Na bazie tej miary można zdefiniować klasyczny współczynnik asymetrii:

$$AS_K(X) = \frac{M_3(X)}{\sqrt[3]{M_2(X)}}, \quad (1)$$

gdzie  $M_2(X)$  określa wariancję zmiennej losowej  $X$ . Inną popularną miarą skośności jest tzw. współczynnik skośności Pearson'a wyznaczany jako:

$$AS_P(X) = \frac{m_1(X) - \text{Moda}(X)}{\sqrt{M_2(X)}}, \quad (2)$$

gdzie  $m_1(X)$  oraz  $\text{Moda}(X)$  oznaczają odpowiednio moment zwykły rzędu pierwszego oraz dominantę (modę) zmiennej losowej  $X$ . Zarówno  $AS_K(X)$  jak i  $AS_P(X)$  informują nie tylko o kierunku asymetrii, ale także o jej sile.

Ciekawe informacje o charakterze skośności rozkładu zmiennej losowej przekazują kwantylowe mierniki asymetrii. W pracy Hinkley'a z 1975 roku [Hinkley 1975] pojawia się uogólniona miara asymetrii bazująca na wybranym kwantylu zmiennej losowej. Zakładając  $p$ -ty kwantyl zmiennej losowej  $X$  oraz definiując medianę tej zmiennej dla  $p = 0,5$  uogólniona kwantylowa miara asymetrii wyrażona jest formułą:

$$AS_Q(X) = \frac{F^{-1}(1-p) - 2F^{-1}(0,5) + F^{-1}(p)}{F^{-1}(1-p) - F^{-1}(p)}, \quad (3)$$

gdzie  $F^{-1}(p)$  oznacza funkcję kwantylową zmiennej losowej  $X$ .

Kwantylowa miara skośności dana wzorem (3) jest uogólnieniem wprowadzonego przez Bowley'a [Bowley 1920] kwantylowego wskaźnika skośności zadanego wzorem:

$$AS_B(X) = \frac{F^{-1}(0,75) - 2F^{-1}(0,5) + F^{-1}(0,25)}{F^{-1}(0,75) - F^{-1}(0,25)}. \quad (4)$$

Zjawisko asymetrii ma swoje odzwierciedlenie w grubości lewego oraz prawego ogona rozkładu. W przypadku asymetrii prawostronnej istnieje wysokie prawdopodobieństwo występowania obserwacji o wartościach istotnie wyższych od poziomu przeciętnego (określonego jako średnia arytmetyczna lub mediana), natomiast w przypadku asymetrii lewostronnej – o wartościach istotnie niższych.



Zjawisko grubych ogonów związane jest z występowaniem w zbiorze danych obserwacji odstających (outliers), które znacząco wpływają na wyniki prowadzonych analiz. Rozwiązanie kwestii pomiaru asymetrii w przypadku zbiorów danych posiadających obserwacje odstające uzyskuje się za pomocą miary MedCouple, wprowadzonej przez Brys'a i in. [Brys i in. 2004].

Dla uporządkowanego zbioru realizacji zmiennej losowej  $X$  niech dany będzie szereg  $x_1 < x_2 < \dots < x_n$ . Niech ponadto zachodzi  $x_i \neq x_j, x_i < x_j$  oraz niech określona będzie pewna funkcja  $h$  jako:

$$h(x_i, x_j) = \frac{x_j - F^{-1}(0,5) + x_i}{x_j - x_i}. \quad (5)$$

W przypadku, gdy  $x_i = x_j = F^{-1}(0,5)$  funkcja  $h$ , określona jako:

$$h(x_i, x_j) = \begin{cases} -1, & i < j \\ 0, & i = j \\ 1, & i > j \end{cases}. \quad (6)$$

Ostatecznie miarę MedCouple definiuje się jako:

$$MC(X) = \text{med}_{x_i \leq F^{-1}(0,5) \leq x_j} h(x_i, x_j), \quad (7)$$

gdzie  $\text{med}(h)$  oznacza medianę zbioru wartości funkcji  $h$ .

### Wprowadzanie skośności do rozkładów prawdopodobieństwa

Biorąc pod uwagę klasyczne podejście do analizy zjawisk obserwowanych na szeroko rozumianym rynku finansowym zdecydowana większość modeli teoretycznych bazuje na symetrycznych rozkładach prawdopodobieństwa (np. rozkład normalny o określonych parametrach czy też rozkład t-Studenta o określonej liczbie stopni swobody). Niemniej jednak możliwa jest transformacja rozkładu symetrycznego w rozkład skośny celem uzyskania odpowiedniego dopasowania do danych [Piontek 2007].

Niech zatem  $g(x)$  oznacza funkcję gęstości zmiennej losowej  $X$  opisaną pewnym ciągłym rozkładem symetrycznym. Niech dodatkowo dane będą funkcje normujące  $u_1(\xi)$  oraz  $u(\xi)$ . Skośny rozkład  $f_\xi(x)$  opisuje następująca formuła:

$$f_\xi(x) = \frac{2[g(\frac{x}{u_1(\xi)})I_{x \geq 0} + g(\frac{x}{u_2(\xi)})I_{x < 0}]}{u_1(\xi) + u_2(\xi)}, \quad (8)$$

gdzie  $I_x$  oznacza funkcję wskaźnikową. Jeśli  $f_\xi(x)$  definiuje skośny rozkład prawdopodobieństwa, jako zmienne normujące  $u$  przyjmuje się najczęściej następujące pary:

$$u_I = \begin{cases} u_1(\xi) = \xi \\ u_2(\xi) = \frac{1}{\xi} \end{cases} \quad (9)$$

lub

$$u_{II} = \begin{cases} u_1(\xi) = 1 - \xi \\ u_2(\xi) = 1 + \xi \end{cases}. \quad (10)$$

Funkcja normująca  $u_I$  posiada parametr skośności  $\xi \in (0, +\infty)$ . Dla  $\xi \in (0,1)$  rozkład jest lewostronnie skośny, natomiast dla  $\xi \in (1, +\infty)$  rozkład jest prawostronnie skośny. Z kolei funkcja normująca  $u_{II}$  posiada parametr skośności  $\xi \in (-1,1)$ . Dla  $\xi \in (-1,0)$  otrzymany rozkład jest lewostronnie skośny, natomiast dla  $\xi \in (0,1)$  rozkład posiada prawostronną asymetrię. Rozkłady symetryczne uzyskuje się dla  $\xi = 1$  (funkcja  $u_I$ ) oraz dla  $\xi = 0$  (funkcja  $u_{II}$ ).

Rozważając praktyczną ocenę asymetrii empirycznego rozkładu prawdopodobieństwa wśród bardziej popularnych rozkładów symetrycznych przekształcanych na rozkłady asymetryczne wyróżnić można skośny rozkład normalny oraz skośny rozkład t-Studenta, dane następującymi wzorami [Piontek 2005]:

- skośny rozkład normalny:

$$f_{s-norm}(x) = \frac{2}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\alpha \frac{x-\mu}{\sigma}\right), \quad (11)$$

gdzie  $X$  jest ciągłą zmienną losową o skośnym rozkładzie normalnym opisanym parametrami położenia  $\mu$ , skali  $\sigma$  oraz kształtu  $\alpha$ , a dodatkowo  $\varphi$  oraz  $\Phi$  oznaczają odpowiednio funkcję gęstości oraz funkcję dystrybuanty rozkładu normalnego standardowego.

- skośny rozkład t-Studenta:

$$f_{s-t}(x; v) = \frac{2}{\xi + \frac{1}{\xi}} \left\{ g(\xi(ax+b); v) I_{x < -\frac{b}{a}} + g\left(\frac{ax+b}{\xi}; v\right) I_{x \geq -\frac{b}{a}} \right\}, \quad (12)$$

gdzie  $X$  jest ciągłą zmienną losową,  $\xi$  jest parametrem skośności,  $g$  jest funkcją gęstości standardowego rozkładu t-Studenta o  $v$  stopniach swobody, natomiast parametry  $a$  oraz  $b$  określają wzory:

$$a = \frac{\Gamma\left(\frac{v-1}{2}\right) \sqrt{v-2}}{\sqrt{\pi} \Gamma\left(\frac{v}{2}\right)} \left(\xi - \frac{1}{\xi}\right) \quad (13)$$

oraz

$$b^2 = \left(\xi^2 + \frac{1}{\xi^2} - 1\right) - a^2, \quad (14)$$

gdzie  $\Gamma$  jest funkcją gamma.

### Kwantylowe miary ryzyka inwestycyjnego

Problem pomiaru ryzyka inwestycyjnego jest zagadnieniem nad wyraz skomplikowanym, zależnym w dużej mierze od przedmiotu inwestycji. W artykule skoncentrowano się przede wszystkim na ocenie ryzyka inwestycji w aktywa finansowe notowane na rynku regulowanym, dla których pomiar dokonywany jest w oparciu o stopy zwrotu. Generalizując definicję ryzyka należy skupić się nie tylko na potencjalnych stratach w stosunku do wartości oczekiwanej, ale także na potencjalnych zyskach. Taki sposób oceny wynika między innymi z faktu, że rozkład zmiennej, której ryzyko dotyczy, zazwyczaj nie jest rozkładem symetrycznym, a zatem stosunek prawdopodobieństwa zysków do strat nie jest

jednostkowy. Na szczególną uwagę zasługują kwantylowe miary ryzyka, które związane są bezpośrednio z realizacją stopy zwrotu na poziomie znacznie różniącym się od oczekiwanego. Wśród kwantylowych miar ryzyka szczególną rolę odgrywa wartość zagrożona VaR, definiowana jako  $\alpha$ -kwantyl rozkładu stopy zwrotu<sup>1</sup> i wskazująca liczbowo wielkość potencjalnej straty, jaka może wynikać z podjętej inwestycji w zadanym horyzoncie czasowym i przy ustalonym poziomie tolerancji  $1 - \alpha$ . VaR posiada jednak pewną wadę – nie jest miarą koherentną. Miernik nie spełnia aksjomatu subaddytywności, tj. założenia, że ryzyko całkowite podjętej inwestycji jest nie większe niż suma ryzyk indywidualnych, składających się na tą inwestycję [Artzner i in. 1999]. Wady tej pozbawione są dwie inne kwantylowe miary ryzyka [Kou i in. 2014]:

- warunkowa wartość zagrożona w zensie wartości oczekiwanej (Expected Shortfall), definiowana jako wartość oczekiwana ponad poziom VaR:

$$ES_{\alpha} = CVaR_{\alpha} = E(r_t - VaR_{\alpha} | r_t > VaR_{\alpha}), \quad (15)$$

- warunkowa wartość zagrożona w sensie mediany (Median Shortfall), definiowana jako mediana realizacji stopy zwrotu ponad poziom VaR:

$$MS_{\alpha} = Median(r_t - VaR_{\alpha} | r_t > VaR_{\alpha}), \quad (16)$$

gdzie  $r_t$  oznacza stopę zwrotu w okresie  $t$ .

Oba mierniki spełniają aksjomaty koherentnej miary ryzyka, a dodatkowo miara Median Shortfall jest odpornym kwantylowym miernikiem ryzyka, w związku z tym może być z powodzeniem wykorzystywana do oceny ryzyka realizacji stopy zwrotu ponad poziom VaR nawet w sytuacji występowania obserwacji ekstremalnych (outliers).

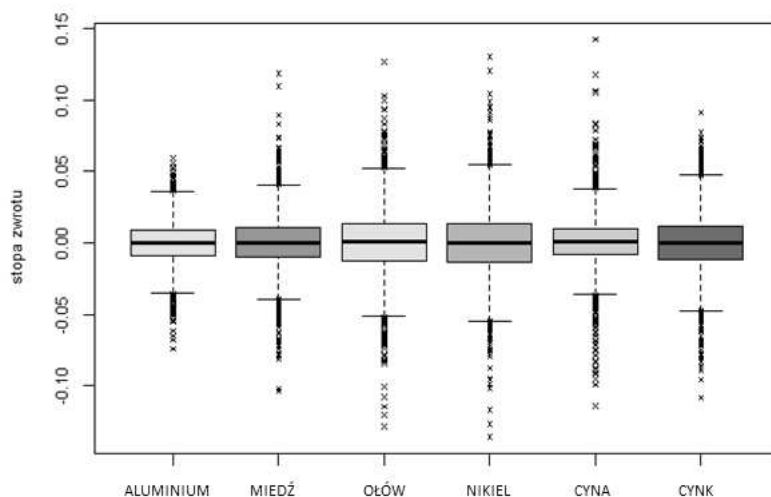
## ANALIZA EMPIRYCZNA

W części empirycznej pracy przedstawiono problem asymetrii w danych oraz jej wpływu na ocenę ryzyka za pomocą miar kwantylowych. Przedmiotem badania są inwestycje realizowane na rynku metali nieżelaznych: aluminium, miedź, ołów, nikiel, cyna oraz cynk. Na podstawie dziennych kursów spot dla badanych metali notowanych na London Metal Exchange w okresie styczeń 2006 – czerwiec 2015 wyznaczonoienne logarytmiczne stopy zwrotu, które były przedmiotem analizy. Zaproponowano porównanie rozkładów symetrycznych: normalnego oraz t-Studenta z wyznaczonymi na ich podstawie analogicznymi rozkładami skośnymi. Wśród mierników kwantylowych wykorzystano VaR oraz ES dla kwantyli na poziomie 0,01, 0,05, 0,95 oraz 0,99. Graiczną prezentację rozkładów przedstawiono na wykresie pudełkowym.

---

<sup>1</sup> Jest to jedna z wielu definicji miary prezentowanych w literaturze (przyp. autora).

Rysunek 1. Wykres pudełkowy empirycznych rozkładów stóp zwrotu analizowanych metali



Źródło: obliczenia własne

Wykres nie wskazuje jednoznacznie na występowanie asymetrii w danych, ale na jego podstawie można zidentyfikować istotną ilość realizacji stopy zwrotu zarówno w lewym jak i w prawym ogonie rozkładu. Podstawowe statystyki opisowe przedstawia tabela 1.

Tabela 1. Statystyki opisowe

Miara	Aluminium	Miedź	Ołów	Nikiel	Cyna	Cynk
średnia	-0,00013	0,00011	0,00021	-0,00006	0,00031	0,00002
odchylenie standardowe	0,01515	0,01959	0,02340	0,02446	0,01991	0,02163
kurtoza	1,39506	3,34158	2,65182	2,51252	5,03549	1,86027
minimum	-0,07437	-0,10400	-0,12850	-0,13605	-0,11435	-0,10832
maksimum	0,05913	0,11880	0,12675	0,13060	0,14253	0,09135

Źródło: obliczenia własne

Statystyki podstawowe wskazują, że w badanym okresie inwestycje w aluminium oraz nikiel generowały średniokresowe straty. Ponadto inwestycje w nikiel cechowały się najwyższym poziomem ryzyka. Z kolei najwyższy oczekiwany zysk zaobserwowano dla inwestycji w cynę, przy relatywnie niskim ryzyku (porównując z pozostałymi aktywami). Empiryczne rozkłady stóp zwrotu dla każdego z metali wykazują wysoki stopień leptokurtozy. W tabeli 2 zaprezentowano wyniki oceny skośności tych rozkładów.

Tabela 2. Miary skośności

Miara	Aluminium	Miedź	Ołów	Nikiel	Cyna	Cynk
MC	0,04014	0,01102	-0,01107	0,01807	-0,02695	-0,00768
AS <sub>K</sub>	-0,15932	-0,03863	-0,20146	-0,01393	-0,13275	-0,18044
AS <sub>P</sub>	-0,00827	0,00557	0,00877	-0,00254	0,01550	0,00076
AS <sub>Q</sub>	0,04876	0,04145	-0,03028	0,03331	-0,00507	0,00401
AS <sub>Q=0,05</sub>	0,03265	0,00152	-0,04002	-0,00693	-0,08454	0,02161
AS <sub>Q=0,01</sub>	-0,05618	0,00586	-0,06783	0,00664	-0,04071	-0,03664

Źródło: obliczenia własne

Wyniki oceny asymetrii w empirycznych rozkładach wykazują różnice ze względu na przyjęty miernik. Miary MC oraz kwantylowy AS<sub>Q</sub> przyjmują odmienne (dodatnie) co do kierunku wartości niż pozostałe mierniki dla niektórych z metali. Przykładowo, rozkład stopy zwrotu cen aluminium w ocenie MC oraz AS<sub>Q</sub> jest rozkładem prawostronnie skośnym, z kolei według miar AS<sub>K</sub> oraz AS<sub>P</sub> jest to rozkład o lewostronnej asymetrii. Należy zwrócić uwagę, że w przypadku różnic co do znaku wartości mierników kwantylowych (w tym MC) częściej przyjmują wartości oscylujące bliżej zera. Może to być podyktowane faktem wpływu obserwacji odstających i ekstremalnych na wartości miar klasycznych.

Do oceny poziomu ryzyka inwestycji w analizowane metale wykorzystano wartość zagrożoną VaR oraz miarę ES. Oszacowano wartości miar dla rozkładu empirycznego, skośnego normalnego oraz skośnego t-Studenta dla kwantyla na poziomie 0,01, 0,05, 0,95 oraz 0,99. Dodatkowo w tabeli zamieszczono oszacowania parametru skośności  $\xi$  (zgodnie z funkcją normującą  $u_I$ ) dla każdego z prezentowanych rozkładów teoretycznych. Wyniki przedstawiono w tabelach 3-4.

Tabela 3. Oszacowanie VaR

VaR	Aluminium	Miedź	Ołów	Nikiel	Cyna	Cynk
Rozkład empiryczny						
VaR <sub>0,01</sub>	-0,04346	-0,05599	-0,06650	-0,06284	-0,05605	-0,06118
VaR <sub>0,05</sub>	-0,02390	-0,03089	-0,03865	-0,03997	-0,03241	-0,03455
VaR <sub>0,95</sub>	0,02471	0,03099	0,03694	0,03880	0,02915	0,03608
VaR <sub>0,99</sub>	0,03810	0,05665	0,05929	0,06305	0,05353	0,05685
Rozkład skośny normalny						
$\xi$	0,94217	0,95721	0,96338	0,97626	0,98311	0,99036
VaR <sub>0,01</sub>	-0,03537	-0,04544	-0,05422	-0,05696	-0,04599	-0,05030
VaR <sub>0,05</sub>	-0,02504	-0,03210	-0,03828	-0,04029	-0,03343	-0,03556
VaR <sub>0,95</sub>	0,02366	0,02809	0,03521	0,03627	0,02739	0,03542
VaR <sub>0,99</sub>	0,03624	0,04527	0,05631	0,06022	0,04992	0,05127

Rozkład skośny t-Studenta						
$\xi$	0,92132	0,93196	0,94188	0,95173	0,96419	0,98128
VaR <sub>0,01</sub>	-0,03977	-0,05125	-0,05744	-0,05871	-0,04723	-0,05437
VaR <sub>0,05</sub>	-0,02419	-0,03176	-0,03831	-0,04014	-0,03275	-0,03497
VaR <sub>0,95</sub>	0,02391	0,02911	0,03572	0,03724	0,02843	0,03579
VaR <sub>0,99</sub>	0,03786	0,04971	0,05727	0,06182	0,05019	0,05217

Źródło: obliczenia własne

Tabela 4. Oszacowanie ES

ES	Aluminium	Miedź	Ołów	Nikiel	Cyna	Cynk
Rozkład empiryczny						
ES <sub>0,01</sub>	-0,05179	-0,07126	-0,08290	-0,08282	-0,07631	-0,07669
ES <sub>0,05</sub>	-0,03459	-0,04590	-0,05539	-0,05514	-0,04878	-0,05043
ES <sub>0,95</sub>	0,03259	0,04489	0,05169	0,05511	0,04469	0,04809
ES <sub>0,99</sub>	0,04417	0,06831	0,07610	0,08170	0,07370	0,06573
Rozkład skośny normalny						
$\xi$	0,94217	0,95721	0,96338	0,97626	0,98311	0,99036
ES <sub>0,01</sub>	-0,04050	-0,05208	-0,06215	-0,06525	-0,05273	-0,05762
ES <sub>0,05</sub>	-0,03137	-0,04028	-0,04806	-0,05051	-0,04074	-0,04459
ES <sub>0,95</sub>	0,02982	0,04173	0,04896	0,05320	0,04198	0,04612
ES <sub>0,99</sub>	0,04276	0,06018	0,06944	0,07532	0,06866	0,06320
Rozkład skośny t-Studenta						
$\xi$	0,92132	0,93196	0,94188	0,95173	0,96419	0,98128
ES <sub>0,01</sub>	-0,04281	-0,05941	-0,07412	-0,07219	-0,06432	-0,06220
ES <sub>0,05</sub>	-0,03219	-0,04186	-0,05122	-0,05184	-0,04313	-0,04616
ES <sub>0,95</sub>	0,03019	0,04220	0,04917	0,05388	0,04313	0,04699
ES <sub>0,99</sub>	0,04320	0,06430	0,07388	0,07813	0,07093	0,06412

Źródło: obliczenia własne

Wyniki zamieszczone w tabelach 3-4 informują, że oszacowanie parametru skośności dla klasycznie symetrycznych rozkładów prawdopodobieństwa wskazuje na lewostronną asymetrię empirycznych rozkładów. Asymetria jest słaba, niemniej jednak występuje. Widoczne jest to już przy porównaniu miar dla kwantyli symetrycznych parami, tj. 0,01-0,99 oraz 0,05-0,95. Biorąc pod uwagę wykorzystanie rozkładów skośnych w estymacji VaR i ES zaobserwowano, że oszacowania bliższe rzeczywistym uzyskano dla rozkładu skośnego t-Studenta. Rozkład ten, w porównaniu z rozkładem normalnym posiada w wersji

standardowej grubszy ogon, co przekłada się na jego wersję skośną. Zatem zasadna jest estymacja miar ryzyka z wykorzystaniem skośnego rozkładu t-Studenta.

## PODSUMOWANIE

W prezentowanej pracy podjęto dyskusję na temat możliwości oceny asymetrii rozkładu empirycznego przy pomocy różnych mierników oraz poruszono problem wpływu asymetrii na oszacowanie wybranych kwantylowych mierników ryzyka. Rozważania teoretyczne zobrazowano na przykładzie inwestycji na rynku metali nieżelaznych. Ocena stopnia asymetrii empirycznych rozkładów stóp zwrotu analizowanych metali wykazała różnice w zależności od przyjętej miaryprzede wszystkim w sensie kierunku asymetrii. Wynik ten można tłumaczyć występowaniem obserwacji ekstremalnych w przestrzeni realizacji stóp zwrotu oraz rozbieżnością rozkładu empirycznego z rozkładem normalnym (rozkłady empiryczne cechowały się grubymi ogonami). Do oceny ryzyka wykorzystano dwie popularne miary: wartość zagrożoną oraz warunkową wartość zagrożoną. Jako rozkłady teoretyczne dla stóp zwrotu zaproponowano skośny rozkład normalny oraz skośny rozkład t-Studenta. Wyniki analizy wskazują, że bardziej rzeczywiste oszacowania miar ryzyka uzyskano przy wykorzystaniu skośnego rozkładu t-Studenta. Jest to uzasadnione, gdyż rozkład ten ma własność uwzględniania występowania w danych obserwacji ekstremalnych. Reasumując należy podkreślić konieczność uwzględniania momentów wyższych rzędów w analizie i ocenie ryzyka inwestycyjnego, gdyż pominięcie tego faktu może skutkować niewłaściwymi prognozami oczekiwanych zysków lub strat.

## BIBLIOGRAFIA

- Aczel A. D. (2000) Statystyka w zarządzaniu. Wydawnictwo Naukowe PWN, Warszawa, 34-35.
- Artzner P., Delbaen F., Eber J.-M., Heath D. (1999) Coherent Measures of Risk. *Mathematical Finance*, 9 (3), 203-228.
- Bowley A. L. (1920) *Elements of Statistics*, Charles Scribner's Sons, New York.
- Brys G., Hubert M., Struyf A. (2004) A Robust Measure of Skewness. *Journal of Computational and Graphical Statistics*, 13 (4), 996-1017.
- Hinkley D. V. (1975) On Power Transformation to Symmetry. *Biometrika*, 62, 101-111.
- Kou S., Peng X. (2014) Expected Shortfall or Median Shortfall. *Journal of Financial Engineering*, 1 (1), 1-6.
- Piontek K. (2005) Modelowanie własności szeregów stóp zwrotu – skośność rozkładów. *Prace Naukowe Akademii Ekonomicznej we Wrocławiu 1096, Ekonometria 15*, 297-308.
- Piontek K. (2007) Pomiar i testowanie skośności rozkładów stóp zwrotu instrumentów finansowych. *Prace Naukowe Akademii Ekonomicznej we Wrocławiu 1096, Taksonomia*, 122-130.

---

**SKEWNESS OF THE DISTRIBUTION AND ESTIMATION  
OF QUANTILE RISK MEASURES – THE CASE OF METALS  
MARKET**

**Abstract:** Investment risk measurement requires specific statistical tools which take into account anomalies observed in empirical distributions of returns. Classical models used for modelling risk are based on gaussian approach and do not include asymmetry in data, which is significantly related to extreme observations. These observations affect the thickness of both right and left tails of the empirical distributions. In this paper the influence of skewness observed in empirical probability distributions on the assessment of extreme risk is examined. The area of research is the metals market within the period including economic crisis. The analysis contains some selected quantile risk measures and their estimation using chosen theoretical distributions.

**Keywords:** skewness, risk measurement, Value-at-Risk, extreme risk, heavy tails



## STRUKTURA WIELKOŚCIOWA PRZEDSIĘBIORSTW W POLSCE

**Sławomir Lisek**

Wydział Rolniczo - Ekonomiczny  
Uniwersytet Rolniczy w Krakowie  
e-mail: slisek@ar.krakow.pl

**Streszczenie:** Przedsiębiorstwa są jednym z głównych podmiotów gospodarki krajowej. W Polsce przeważają przedsiębiorstwa zatrudniające nie więcej niż dziesięć osób (96%), w których pracuje 39%, a których przychody satniają 22% ogółu. Celem artykułu jest analiza struktury wielkości przedsiębiorstw w Polsce mierzonej udziałem poszczególnych grup w ilości przedsiębiorstw, w zatrudnieniu i w przychodach ogółem. Wykorzystano dane ze strony internetowej GUS. Badania przeprowadzono wykorzystując metodę Chomątowskiego i Sokołowskiego. Zróżnicowanie międzywojewódzkie struktury wielkościowej przedsiębiorstw jest umiarkowane.

**Słowa kluczowe:** przedsiębiorstwo, wielkość przedsiębiorstwa, struktura ekonomiczna

### WSTĘP

Przedsiębiorstwa są zasadniczym elementem gospodarki narodowej. Są one głównym źródłem strumieni dochodów dla rodzin, jednostek lokalnych, czy też krajów. Celem przedsiębiorstwa jest głównie wygospodarowywanie zysku [Sudoł 1999], który osiągają poprzez wypełnianie swoich funkcji. Funkcja przedsiębiorstwa to zaspokajanie potrzeb osób, instytucji, zbiorowości, wynikających z przedmiotu działalności firmy (tamże). Coraz częściej pada także postulat społecznej odpowiedzialności przedsiębiorstwa. Ogólnie dzieli się ją na wewnętrzną i zewnętrzną. Wewnętrzna to uwzględnienie w zarządzaniu przedsiębiorstwem uzasadnionych postulatów pracowników. Natomiast zewnętrzna to dbałość o potrzeby otoczenia bliższego lub dalszego, taka jak: ochrona środowiska, stwarzanie miejsc pracy, przyczynienie się do rozwoju społeczno – ekonomicznego danego regionu (tamże). Co do istnienia społecznej

odpowiedzialności przedsiębiorstwa zdania są podzielone, niemniej jednak, nawet gdyby uznać, że ona nie istnieje, to przedsiębiorstwa swoją działalnością automatycznie przynoszą korzyści ekonomiczne otoczeniu. Generalnie można zaryzykować postawienie hipotezy, że im przedsiębiorstwo jest większe, tym więcej przynosi korzyści otoczeniu, natomiast małe firmy skupione są głównie na zyskach właścicieli. Korzyści z prowadzenia działalności przez przedsiębiorstwa nie są rozłożone równomiernie. Teoria Simona Kuzneta głosi, że skala nierówności jest większa przy niższych niż przy wyższych dochodach. W miarę wzrostu dochodów, nierówności zaczynają się zacierać [Jasiński 2013]. Dlatego też godną uwagi jest zbadanie struktury przedsiębiorstw ze względu na ich wielkość mierzonej ilości przedsiębiorstw, wielkością zatrudnienia i przychodami ze sprzedaży.

Celem niniejszego artykułu jest analiza struktury wielkości przedsiębiorstw w Polsce określanej udziałem poszczególnych grup przedsiębiorstw w ilości przedsiębiorstw ogółem, w zatrudnieniu ogółem i w przychodach ze sprzedaży, a także dokonanie oceny zróżnicowania tych struktur w województwach w 2015r.

## MATERIAŁ I METODA ANALIZY

Źródłem badań w niniejszym artykule jest publikacja Podstawowe dane o przedsiębiorstwach według klas wielkości, zawarta na stronie internetowej GUS<sup>1</sup>. Za miarę zróżnicowania struktur w niniejszej pracy przyjmuje się [Kukuła 1996]:

$$v_{\alpha,\beta} = \frac{\sum_{i=1}^k |\alpha_i - \beta_i|}{2}, \quad (1)$$

gdzie:  $\alpha_i$ - i – ta składowa struktury  $\alpha$ ;  $\beta_i$ - i – ta składowa struktury  $\beta$ ; k – liczba składowych struktury

Grupowania województw w podzbiory podobne ze względu na omawiane struktury dokonano metodą eliminacji wektorów Chomątowskiego i Sokołowskiego [Chomątowski, Sokołowski 1978]. Wartość krytyczna została ustalona, jako [Kukuła 1996]:

$$\varepsilon = \frac{2 \sum_{j=1}^r \sum_{p>j} v_{jp}}{r(r-1)}, \quad (2)$$

gdzie:  $v_{j,p}$  – miara zróżnicowania obiektów j i p; r – liczba obiektów.

---

<sup>1</sup> <http://stat.gov.pl/obszry-tematyczne/podmioty-gospodarcze-wyniki-finansowe/przedsiębiorstwa-niefinansowe/dzialalnosc-przedsiębiorstw-niefinansowych-w-2015-r-,2,11.html> - dostęp 03.06.2017r.

## WYNIKI BADAŃ

Podstawowe dane dotyczące ogółu przedsiębiorstw w Polsce, w podziale na grupy wielkościowe, zaprezentowano w tabeli 1.

Tabela 1. Struktura wielkościowa przedsiębiorstw w Polsce według wyszczególnionych kryteriów w 2015 roku

Lp.	Liczba pracowników	Struktura wg liczby	Struktura wg zatrudnienia	Struktura wg przychodów
1	0-9	0,960	0,390	0,220
2	10-49	0,030	0,127	0,142
3	50-249	0,008	0,174	0,197
4	250 i więcej	0,002	0,309	0,442
	Razem	1,000	1,000	1,000

Źródło: opracowanie własne

W Polsce zdecydowanie dominują najmniejsze przedsiębiorstwa, zatrudniające nie więcej niż dziewięciu pracowników. Jest ich ponad 1,8 mln (96%), generują 3,7 mln (39%) miejsc pracy i 895.524 mln PLN przychodu (22%). Na jedną firmę z tej grupy, przypada więc przeciętnie dwóch pracowników i 487 tys. PLN przychodu. W grupie przedsiębiorstw o zatrudnieniu od 10 do 49 osób występuje 56,7 tys. firm (3%), zatrudniających 1,2 mln pracowników (12,7%) i generujących 577.808 mln PLN przychodu (14%). Jedna firma z tej grupy daje więc średnio 21 miejsc pracy i pozwala wygenerować średnio 10.188tys. PLN przychodu. Przedsiębiorstwa większe, zatrudniające od 50 do 249 pracowników, których liczba wynosi 15,6 tys. (0,8%), dają średnio 104 miejsca pracy (17%) i wytwarzają 51,3 mln PLN przychodu (20%). Największych firm jest 3,4 tys. (0,2%), średnio na jedno przypada w nich 846 pracowników (30,9%) i generują średnio 524,86mln PLN przychodu na firmę (44%).

Należy, więc powiedzieć, że większe przedsiębiorstwa generują znacznie więcej korzyści dla otoczenia. Dlatego też im większy jest udział w strukturze dużych przedsiębiorstw, tym lepiej, oczywiście jeśli osiąga się to dzięki powstawaniu dużych firm, a nie likwidacji małych. Jednak duże przedsiębiorstwa wiążą się również z większym ryzykiem. Upadek małej firmy nie przynosi właściwie otoczeniu odczuwalnych konsekwencji. Natomiast w razie upadłości dużego przedsiębiorstwa, lawinowo rosną negatywne konsekwencje dla otoczenia, ponieważ likwidowana jest natychmiast duża ilość miejsc pracy, gwałtownie maleją przychody w danym regionie, co za tym idzie zmniejsza się produkt brutto regionu i maleje wartość dodana.

W tabeli 2 przedstawiono porównanie struktury wielkości przedsiębiorstw Polski ogółem wg wybranych kryteriów z odpowiednimi strukturami poszczególnych województw mierzone miarą zróżnicowania (1).

Tabela 2. Miara zróżnicowania pomiędzy strukturą wielkością przedsiębiorstw Polski ogółem, a strukturą wielkością poszczególnych województw w 2015 roku

Województwo	Struktura wg liczby	Struktura wg zatrudnienia	Struktura wg przychodów
Dolnośląskie	0,004	0,033	0,021
Kujawsko-Pomorskie	0,002	0,086	0,077
Lubelskie	0,004	0,110	0,154
Lubuskie	0,000	0,108	0,123
Łódzkie	0,001	0,038	0,050
Małopolskie	0,002	0,053	0,057
Mazowieckie	0,005	0,148	0,076
Opolskie	0,005	0,117	0,114
Podkarpackie	0,010	0,071	0,010
Podlaskie	0,001	0,123	0,161
Pomorskie	0,002	0,106	0,020
Śląskie	0,007	0,022	0,015
Świętokrzyskie	0,002	0,110	0,070
Warmińsko-Mazurskie	0,001	0,103	0,148
Wielkopolskie	0,001	0,026	0,061
Zachodniopomorskie	0,008	0,142	0,221

Źródło: opracowanie własne

Pod względem ilości przedsiębiorstw, zróżnicowanie pomiędzy całym krajem a poszczególnymi województwami jest znikome (tabela 2). Maksymalną wartość osiąga 0,010, więc struktura jest w zasadzie jednorodna. Zróżnicowanie pod względem ilości przedsiębiorstw nie będzie więc dalej analizowane.

Większe zróżnicowanie występuje w strukturze wielkościowej mierzonej proporcją zatrudnienia, gdzie najbardziej podobną strukturą do ogólnopolskiej charakteryzuje się województwo śląskie, którego miara zróżnicowania od reszty kraju wynosi 0,022, najmniej podobną - mazowieckie (miara zróżnicowania 0,148) oraz zachodniopomorskie (miara zróżnicowania 0,142).

Omawiając strukturę mierzoną udziałem w przychodach ze sprzedaży, to najmniejsze zróżnicowanie pomiędzy Polską ogółem wykazuje województwo podkarpackie (miara zróżnicowania 0,010), natomiast największe zachodniopomorskie (miara zróżnicowania 0,221).

Grupując województwa w podzbiory podobnych pod względem struktury wielkościowej mierzonej udziałem liczby zatrudnionych, otrzymuje się cztery grupy województw podobnych (tabela 3).

Tabela 3. Grupy województw podobnych pod względem struktury wielkościowej mierzonej udziałem w zatrudnieniu w 2015 roku, z podanymi średnimi udziałami w każdej grupie przedsiębiorstw

Grupa	Województwa	Wielkość			
		0-9	10-49	50-249	250 i więcej
I	Kujawsko – Pomorskie, Lubelskie, Lubuskie, Łódzkie, Małopolskie, Opolskie, Podkarpackie, Podlaskie, Pomorskie, Świętokrzyskie, Warmińsko – Mazurskie	0,436	0,156	0,194	0,219
II	Dolnośląskie, Śląskie, Wielkopolskie	0,376	0,123	0,169	0,332
III	Mazowieckie	0,314	0,088	0,141	0,457
IV	Zachodniopomorskie	0,511	0,141	0,181	0,167

Źródło: opracowanie własne

Maksymalne zróżnicowanie struktur wyznaczanych proporcją zatrudnienia wynosi wg miary zróżnicowania 0,290 i dotyczy województw mazowieckiego i zachodniopomorskiego. Najbardziej podobne zaś są do siebie województwa pomorskie i lubuskie (miara zróżnicowania 0,007).

Zbiorowość województw grupuje się w dwa zbiory o znacznie różniącej się liczebności, ponadto występują dwa elementy izolowane.

Najliczniejsza grupa obejmuje 11 województw. W tej grupie w strukturze zatrudnienia dominują firmy zatrudniające do 9 pracowników. Ich udział w strukturze wynosi 0,436. Udział największych przedsiębiorstw wynosi 0,219. Niewiele mniejszy udział mają firmy o zatrudnieniu 50-249 pracowników 0,194. Dla grupy tej charakterystycznym jest, więc dominujący udział najmniejszych firm w strukturze zatrudnienia.

Mniej liczna grupa obejmuje województwa: dolnośląskie, śląskie i wielkopolskie. W tej grupie średni udział firm najmniejszych (0,376), jest zbliżony do średniego udziału firm największych (0,332). Firmy zatrudniające 50-249 pracowników stanowią 0,169 przedsiębiorstw z tej grupy. Należy, więc powiedzieć, iż w tej grupie znacznie większy udział w zatrudnieniu stanowią największe firmy. Województwa te, są też lepiej rozwinięte gospodarczo od tych z grupy 1.

Występują dwa elementy izolowane: województwo mazowieckie i zachodniopomorskie. Mazowieckie charakteryzuje się najwyższym udziałem największych firm w strukturze zatrudnienia. Jest to też najlepiej rozwinięte gospodarczo województwo w Polsce. Województwo zachodniopomorskie charakteryzuje się zrównoważonym udziałem czterech grup przedsiębiorstw w strukturze zatrudnienia, z niewielką przewagą firm najmniejszych.

Grupując województwa w podzbiory podobnych pod względem struktury wielkościowej mierzonej udziałem przychodów, otrzymuje się następujące grupy:

Tabela 4. Grupy województw podobnych pod względem struktury wielkościowej mierzonej udziałem w przychodach w 2015 roku z podanymi średnimi udziałami w każdej grupie przedsiębiorstw

Lp.	Województwa	Wielkość			
		0-9	10-49	50-249	250 i więcej
1	Dolnośląskie, Kujawsko – Pomorskie, Łódzkie, Małopolskie, Podkarpackie, Pomorskie, Śląskie, Świętokrzyskie,	0,236	0,150	0,204	0,410
2	Lubelskie, Lubuskie, Warmińsko - Mazurskie	0,312	0,174	0,212	0,303
3	Mazowieckie, Wielkopolskie	0,185	0,124	0,181	0,511
4	Opolskie	0,250	0,170	0,252	0,328
5	Podlaskie	0,302	0,188	0,229	0,281
6	Zachodniopomorskie	0,340	0,204	0,235	0,221

Źródło: opracowanie własne

Maksymalne zróżnicowanie struktur wynosi wg miary zróżnicowania 0,297 i dotyczy województw mazowieckiego i zachodniopomorskiego. Najbardziej podobne zaś są do siebie województwa podkarpackie i śląskie (miara zróżnicowania 0,014).

Dzieląc województwa w podzbiory podobne pod względem struktury, otrzymujemy sześć grup, z których I przewyższa liczebnością pozostałe. W najliczniejszej grupie dominuje udział największych firm, natomiast jest ona dosyć znacznie zróżnicowana. Obejmuje ona województwa o średnio dobrej sytuacji ekonomicznej. Grupa II obejmuje trzy województwa, o gorszej sytuacji gospodarczej. Tutaj największy udział w przychodach mają najmniejsze firmy, niewiele niższy największe. Natomiast jest to grupa o dużej jednorodności. Grupę III tworzą dwa województwa: mazowieckie i wielkopolskie, czyli najlepiej rozwinięte gospodarczo. Tutaj udział w przychodach przedsiębiorstw największych przekracza połowę. Elementy izolowane to województwa: opolskie, podlaskie i zachodniopomorskie. W województwie opolskim w strukturze przychodów dominują firmy największe (0,328), następnie podobny udział mają firmy zatrudniające 50-249 pracowników (0,252), oraz najmniejsze (0,250), najmniejszy zaś udział firmy zatrudniające 10-49 osób (0,170). W województwach podlaskim i zachodniopomorskim dominują w strukturze przychodów firmy najmniejsze (0,302 województwo podlaskie i 0,340 zachodniopomorskie). W województwie podlaskim na drugim miejscu są firmy największe (0,281), na końcu zaś zatrudniające 10-49 pracowników (0,188). W województwie zachodniopomorskim udział firm zatrudniających: 10-49 pracowników, 50-249 pracowników i ponad 250 pracowników jest zrównoważony i wynosi niewiele ponad 0,200.

## KONKLUZJE

W Polsce zdecydowanie dominują przedsiębiorstwa najmniejsze, zatrudniające do 9 pracowników. Dają one średnio 2 miejsca pracy na przedsiębiorstwo, generując 487 tys. PLN przychodu. Przedsiębiorstwa duże (zatrudniające 50-249 pracowników) zatrudniają przeciętnie 104 pracowników, dając średni przychód 51,3 mln zł. Natomiast w największych przedsiębiorstwach (zatrudniających powyżej 250 osób) średnio pracuje 846 osób i generują ponad tysiąckrotnie więcej przychodu w przeliczeniu na przedsiębiorstwo niż przedsiębiorstwa zatrudniające poniżej 10 osób. Większe przedsiębiorstwa przynoszą, więc znacznie więcej korzyści otoczeniu, ale wiąże się to z większym ryzykiem dla otoczenia w razie upadłości firmy.

Przeprowadzone badanie wykazało, iż w strukturze ilościowej w Polsce zdecydowanie przeważają firmy najmniejsze. Jednak w strukturze zatrudnienia i generowanych przychodów ich udział jest znacznie niższy.

Miara zróżnicowania pomiędzy województwami nie przekracza nigdy 0,4, co oznacza dosyć dużą jednorodność struktury wielkościowej przedsiębiorstw w Polsce.

Grupowanie województw w podzbiory podobnych pod względem struktury wielkościowej przedsiębiorstw, generalnie pozwala wyciągnąć wniosek, iż w przypadku województw o lepszej sytuacji ekonomicznej, znacznie większy udział w strukturze zarówno zatrudnienia, jak i przychodów mają przedsiębiorstwa największe.

## BIBLIOGRAFIA

- Chomątowski S., Sokołowski A. (1978) Taksonomia struktur. Przegląd Statystyczny, 2, 217-225.
- Jasiński L. J. (2013) Podstawy ekonomii. Oficyna Wydawnicza Politechniki Warszawskiej.
- Kukuła K. (1996) Statystyczne metody analizy struktur ekonomicznych. Wydawnictwo Edukacyjne, Kraków.
- Sudoł S. (1999) Przedsiębiorstwo, podstawy nauki o przedsiębiorstwie. Teorie i praktyka zarządzania. TNOiK, Toruń.

### MAGNITUDE STRUCTURE OF THE ENTERPRISES IN POLAND

**Abstract:** Enterprises are one of the main elements in economy. Magnitude of the enterprise makes influence into benefit to neighbourhood. In the amount structure smallest enterprises prevail. In the employment and revenues structure prevail bigger enterprises. Differences between voivodships are restrained.

**Keywords:** enterprise, magnitude of the enterprise, economics structure



## METODY KLASYFIKACJI W ANALIZIE PORÓWNAWCZEJ ROZWOJU SPOŁECZNO-GOSPODARCZEGO POLSKICH WOJEWÓDZTW W LATACH 2010 I 2014 – WPŁYW PROCEDURY NORMALIZACJI NA WYNIK RANKINGU

**Jan Malchar, Monika Zielińska-Sitkiewicz**

Wydział Zastosowań Informatyki i Matematyki

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

e-mail: janek.malchar@wp.pl, monika\_zielinska\_sitkiewicz@sggw.pl

**Streszczenie:** Celem artykułu jest zweryfikowanie czy wybór metody klasyfikacji oraz procedury normalizacji zmiennych ma wpływ na utworzone rankingi województw charakteryzujące poziom rozwoju społeczno-gospodarczego w latach 2010 oraz 2014. W badaniu wykorzystano Syntetyczny Miernik Rozwoju (SMR) oraz Wskaźnik Względego Poziomu Rozwoju (BZW). Do normalizacji zmiennych zastosowano formuły standaryzacji klasycznej, standaryzacji Webera oraz unitaryzacji.

**Słowa kluczowe:** ranking, normalizacja, syntetyczne mierniki taksonomiczne, SMR, BZW, rozwój społeczno-gospodarczy

### WSTĘP

Rozwój społeczno-gospodarczy ma charakter złożony i na jego poziom wpływają różnorodne czynniki ekonomiczne, społeczne i geograficzno-przyrodnicze. Według Amartyi Sena [Sen 2002], rozwój społeczny można określić jako wzrost wolności ludzi oraz możliwości życia w taki sposób, jaki jest dla nich wartościowy. Wyróżnić tu można kilka najważniejszych determinant, takich jak zdrowie, edukacja, stopień zaspokojenia ludzkich potrzeb, zamożność i jakość życia. Z kolei rozwój gospodarczy jest procesem zmian realizowanych w wielkości produkcji przemysłowej i rolnej, wysokości zatrudnienia, dochodach i w innych obszarach ekonomicznych. Wpływ tutaj mają także: poziom wykształcenia i umiejętności siły roboczej, postęp techniczny oraz doskonalenie jakości wyrobów. Przyczyniają się one do pozytywnych zmian strukturalnych, powodujących rozwój społeczno-gospodarczy badanego obszaru. Jak istotna jest to

problematyka świadczą o tym liczne prace naukowe z tej dziedziny m. in. Szymli [2000, 2005], Strahl [2006], Stec [2011], Szewczuk [2011], Kukuły [2014], Chrzanowskiej, Dejerskiej [2016], Kisielińskiej [2017].

Czynniki charakteryzujące rozwój społeczno-gospodarczy można podzielić na determinanty o charakterze wewnętrznym i zewnętrznym. W skład pierwszej grupy wchodzi zasoby własne regionu, do których należą: zasoby demograficzne, zasoby z zakresu infrastruktury technicznej, inwestycje zagraniczne, zasobność budżetowa, zasoby związane z ekorozwojem i świadomość ekologiczna społeczności, zasoby w sferze gospodarki regionalnej, zasoby dotyczące przestrzeni regionu, a także skuteczność marketingowa przestrzeni [Strahl 2006]. Do czynników egzogenicznych należą m. in. system społeczno-gospodarczy, polityka ekonomiczna, koniunktura gospodarcza i sojusze polityczne [Kudełko 2013].

Celem pracy jest analiza porównawcza rozwoju społeczno-gospodarczego polskich województw z wykorzystaniem różnych metod klasyfikacji, a następnie zweryfikowanie czy wybór formuły normalizacji zmiennych diagnostycznych oraz procedury klasyfikacji miały wpływ na otrzymane wyniki. Jako okres badawczy wybrano lata 2010 oraz 2014, będące czasem transformacji gospodarczej regionów na skutek wdrażania Funduszy Europejskich. Do analizy wykorzystano dane GUS.

Realizując badanie dokonano doboru odpowiednich zmiennych charakteryzujących rozwój społeczno-gospodarczy, a następnie stworzono rankingi województw oparte na dwóch taksonomicznych miernikach – Syntetycznym Mierniku Rozwoju (SMR) oraz Wskaźniku Względego Poziomu Rozwoju (BZW). Zastosowano trzy metody normalizacji: standaryzację klasyczną, standaryzację Webera oraz unitaryzację, a następnie sprawdzono czy wybór formuły miał istotny wpływ na utworzone klasyfikacje. Zbieżność rankingów województw zmierzono współczynnikiem korelacji Spearmana.

## DOBÓR CECH STATYSTYCZNYCH, METODYKA BADAŃ I WYNIKI

Rozwój regionu jest zjawiskiem mierzalnym i w celu pomiaru wymaga wykorzystania różnych, odzwierciedlających wszystkie istotne cechy, zmiennych. Do badań rozwoju społeczno-gospodarczego województw w pierwszym etapie zaproponowano kilkadziesiąt wskaźników spełniających przesłanki merytoryczne i podzielono je na siedem następujących grup: Ludność, Rynek pracy, Przemysł i budownictwo, Rolnictwo, Usługi socjalne, oświatowe i kulturalne, Transport i łączność oraz Zamożność ludności.

W następnym etapie dokonano wyboru zmiennych diagnostycznych sprawdzając różnicowanie i badając skorelowanie poszczególnych zmiennych. Ustalono wartość krytyczną współczynnika korelacji na poziomie 0,49. Uznano, że każdą grupę cech będzie reprezentowała dokładnie jedna zmienna. Wpływ na tę decyzję miał fakt, iż korelacje między zmiennymi z jednej grupy cech były dosyć silne.

Ponadto biorąc pod uwagę merytoryczne znaczenie cechy określono charakter każdej z nich, wyodrębniając stymulanty (S) oraz destymulanty (D). Ostateczny zbiór zmiennych diagnostycznych przyjął następującą postać:

- X<sub>1</sub> - Przyrost naturalny na 1000 osób (S),
- X<sub>2</sub> - Liczba nowo utworzonych miejsc pracy w tys. na 1000 osób (S),
- X<sub>3</sub> - Odsetek osób pracujących w sektorze przemysłowym (S),
- X<sub>4</sub> - Produkcja żywca rzeźnego w przeliczeniu na mięso na 1 ha użytków rolnych [kg] (S),
- X<sub>5</sub> - Liczba studentów szkół wyższych na 10 tys. ludności (S),
- X<sub>6</sub> - Drogi o twardej nawierzchni ulepszonej wyrażone w km na 100 km<sup>2</sup> (S),
- X<sub>7</sub> - Procent osób w gosp. dom. poniżej ustawowej granicy ubóstwa [%] (D).

Ideą Wielowymiarowej Analizy Porównawczej jest skonstruowanie zagregowanego miernika syntetycznego dla poklasyfikowania badanych obiektów ze względu na poziom wielocechowego zjawiska. Prekursorem metodologii porządkowania liniowego był profesor Zdzisław Hellwig. Jako pierwszy zaproponował syntetyczną miarę rozwoju [Hellwig 1968] dla porównania poziomu rozwoju gospodarczego wybranych krajów. Do rozwoju algorytmów konstruowania taksonomicznych mierników dla różnych zastosowań przyczynili się w Polsce m.in. Cieślak [1974], Bartosiewicz [1976], Strahl [1978], Kurkiewicz, Pocięcha, Zajac [1991], Zeliaś, Malina [1997], Kukuła [1986, 2000], Gatnar, Walesiak [2004] i Walesiak [2003,2014]. Istnieje wiele procedur tworzenia syntetycznych mierników, wykorzystujących odpowiednio wybrane zmienne diagnostyczne. Ze względu na ograniczoną objętość artykułu nie opisano szczegółowo użytych w badaniu narzędzi badawczych. Charakterystykę Syntetycznego Miernika Rozwoju (SMR) i Wskaźnika Względego Poziomu Rozwoju (BZW) można znaleźć w pracach M. Stec [2011] oraz A. Mazur, D. Witkowska [2006]. Procedury normalizacji zmiennych zaprezentowano w pracach M. Walesiaka [2014] i M. Zielińskiej-Sitkiewicz [2017]. Standaryzacja klasyczna i Webera skutkują ujednoczeniem wartości wszystkich zmiennych pod względem zmienności, co oznacza wyeliminowanie jej jako podstawy różnicowania obiektów. Zaleca się stosowanie standaryzacji Webera, gdy rozkład empiryczny badanych zmiennych jest silnie asymetryczny. Formuła normalizacyjna - unitaryzacja zerowana ma tę zaletę, że zapewnia znormalizowanym wartościom zmiennych zróżnicowaną zmienność (mierzoną odchyleniem standardowym) i jednocześnie stały rozstęp dla wszystkich zmiennych [Walesiak 2014].

Na wstępie badań przeanalizowano rozkłady zmiennych, bowiem przy wyborze formuły normalizacyjnej powinno się uwzględnić zarówno skale pomiaru zmiennych, jak również charakterystyki rozkładu zmiennych, takie jak średnia arytmetyczna, odchylenie standardowe oraz rozstęp wyznaczony dla znormalizowanych wartości zmiennych [Walesiak 2014]. W przypadku dwóch

zmiennych:  $X_2$  i  $X_6$  stwierdzono silną asymetrię prawostronną w dwóch badanych okresach i występowanie wartości odstających w 2010 roku.

W celu zweryfikowania, czy dobór formuły normalizacyjnej miał wpływ na końcowe rankingi województw, skonstruowano je z wykorzystaniem trzech metod standaryzacji cech diagnostycznych. Ponadto obliczono współczynniki korelacji rang Spearmana celem ustalenia stopnia zbieżności uzyskanych klasyfikacji. zilustrowano przestrzennie otrzymane wyniki na mapach Polski.

### Syntetyczny Miernik Rozwoju (SMR)

W pierwszej kolejności wykonano uporządkowanie województw za pomocą Syntetycznego Miernika Rozwoju. Wyznaczono wzorce rozwoju, a następnie odległości euklidesowe obiektów od wzorców. Uzyskane wartości SMR i pozycje województw w rankingach dla lat 2010 i 2014 zaprezentowano w tabeli 1.

Tabela 1. Wartości Syntetycznego Miernika Rozwoju (SMR) oraz pozycja województwa w rankingu wg różnych metod normalizacji (lata 2010 oraz 2014)

Jednostka terytorialna (województwo)	Rok 2010						Rok 2014					
	standaryzacja klasyczna		standaryzacja Webera		unitaryzacja		standaryzacja klasyczna		standaryzacja Webera		unitaryzacja	
	SMR	pozycja	SMR	pozycja	SMR	pozycja	SMR	pozycja	SMR	pozycja	SMR	pozycja
Dolnośląskie	0,23	9	0,22	7	0,28	9	0,30	7	0,36	6	0,28	7
Kujawsko-pomorskie	0,27	7	0,21	9	0,29	7	0,27	8	0,27	8	0,27	8
Lubelskie	0,00	16	0,00	15	0,00	16	0,10	14	0,10	15	0,10	13
Lubuskie	0,14	11	0,00	16	0,18	11	0,15	12	0,17	13	0,14	12
Łódzkie	0,25	8	0,27	5	0,29	8	0,30	6	0,31	7	0,29	6
Małopolskie	0,48	2	0,50	1	0,52	2	0,42	4	0,43	2	0,43	2
Mazowieckie	0,34	4	0,31	4	0,36	4	0,43	2	0,42	3	0,42	5
Opolskie	0,27	6	0,22	8	0,31	5	0,24	9	0,26	9	0,23	9
Podkarpackie	0,12	13	0,05	12	0,16	12	0,15	11	0,19	10	0,15	11
Podlaskie	0,07	15	0,02	14	0,07	15	0,00	16	0,00	16	0,00	16
Pomorskie	0,32	5	0,23	6	0,31	6	0,43	3	0,41	5	0,43	3
Śląskie	0,44	3	0,44	3	0,46	3	0,42	5	0,41	4	0,42	4
Świętokrzyskie	0,08	14	0,08	10	0,07	14	0,12	13	0,18	11	0,10	14
Warmińsko-mazurskie	0,12	12	0,02	13	0,10	13	0,06	15	0,10	14	0,04	15
Wielkopolskie	0,54	1	0,45	2	0,52	1	0,48	1	0,51	1	0,46	1
Zachodniopomorskie	0,15	10	0,06	11	0,19	10	0,16	10	0,17	12	0,16	10

Źródło: obliczenia własne

Wyniki rankingów SMR i wartości współczynników korelacji rang Spearmana (por. tabele 2 i 3) potwierdzają, że klasyfikacje województw, budowane

z wykorzystaniem standaryzacji klasycznej oraz unitaryzacji, są do siebie najbardziej zbliżone. Użycie standaryzacji Webera nieznacznie zmieniło kolejność województw i zniwelowało wpływ wartości odstających dwóch zmiennych  $X_2$  i  $X_6$  na końcowy rezultat uporządkowania badanych regionów.

Tabela 2. Wartości współczynników korelacji rang Spearmana (2010)<sup>1</sup>

SMR 2010	SK	SW	U
SK	1		
SW	0,897	1	
U	0,994	0,894	1

Źródło: obliczenia własne

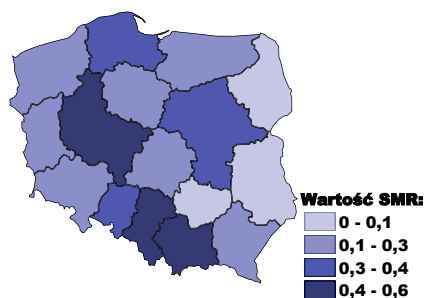
Tabela 3. Wartości współczynników korelacji rang Spearmana (2014)<sup>2</sup>

SMR 2014	SK	SW	U
SK	1		
SW	0,965	1	
U	0,976	0,956	1

Źródło: obliczenia własne

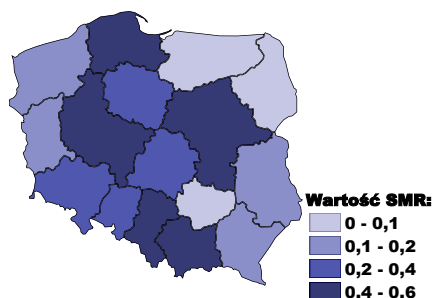
W celu zilustrowania przestrzennego rozmieszczenia wartości miernika SMR utworzono mapy Polski (patrz rysunek 1 i 2). Wykorzystano wyniki rankingu otrzymanego przy użyciu unitaryzacji, która według K. Kukuły i L. Luty [2015] nie niweluje wartości odstających cech diagnostycznych, czyli nie zniekształca rzeczywistego obrazu rozkładu przestrzennego badanego zjawiska złożonego. Wyszczególnione województwa wypełniono tak, aby ciemniejsze kolory wskazywały na wyższe wartości miary syntetycznej, czyli wyższy poziom rozwoju społeczno-gospodarczego.

Rysunek 1. Przestrzenne rozmieszczenie wartości miernika SMR w 2010 r. (unitaryzacja)



Źródło: opracowanie własne

Rysunek 2. Przestrzenne rozmieszczenie wartości miernika SMR w 2014 r. (unitaryzacja)



Źródło: opracowanie własne

<sup>1,2</sup> SK – standaryzacja klasyczna; SW – standaryzacja Webera; U – unitaryzacja.

### Wskaźnik Względego Poziomu Rozwoju (BZW)

Drugą metodą, którą sporządzono klasyfikacje województw pod względem rozwoju społeczno-gospodarczego dla lat 2010 i 2014 był bezwzorcowy Wskaźnik Względego Poziomu Rozwoju BZW. Rankingi otrzymano wykorzystując trzy analizowane formuły normalizacyjne, a wyniki obliczeń BZW i pozycje województw zaprezentowano w tabeli 4.

Tabela 4. Wartości Wskaźnika Względego Poziomu Rozwoju (BZW) oraz pozycja województwa w rankingu wg różnych metod normalizacji (lata 2010 oraz 2014)

Jednostka terytorialna (województwo)	Rok 2010						Rok 2014					
	standaryzacja klasyczna		standaryzacja Webera		unitaryzacja		standaryzacja klasyczna		standaryzacja Webera		unitaryzacja	
	BZW	pozycja	BZW	pozycja	BZW	pozycja	BZW	pozycja	BZW	pozycja	BZW	pozycja
Dolnośląskie	0,47	7	0,48	6	0,48	7	0,51	6	0,50	6	0,51	6
Kujawsko-pomorskie	0,45	9	0,43	9	0,45	9	0,43	8	0,41	9	0,44	8
Lubelskie	0,22	16	0,23	16	0,21	16	0,30	13	0,27	14	0,31	13
Lubuskie	0,41	10	0,35	10	0,43	10	0,35	10	0,34	11	0,36	10
Łódzkie	0,47	8	0,47	7	0,48	8	0,49	7	0,46	7	0,50	7
Małopolskie	0,67	2	0,68	1	0,67	1	0,62	4	0,61	4	0,63	4
Mazowieckie	0,58	4	0,58	4	0,60	4	0,69	1	0,67	1	0,69	1
Opolskie	0,49	6	0,47	8	0,51	5	0,42	9	0,41	8	0,43	9
Podkarpackie	0,36	12	0,33	12	0,38	12	0,35	11	0,34	10	0,35	12
Podlaskie	0,27	15	0,26	15	0,26	14	0,21	16	0,18	16	0,21	16
Pomorskie	0,51	5	0,48	5	0,50	6	0,61	5	0,58	5	0,61	5
Śląskie	0,64	3	0,66	2	0,65	3	0,64	2	0,63	2	0,65	2
Świętokrzyskie	0,27	14	0,28	14	0,26	15	0,29	14	0,29	13	0,29	14
Warmińsko-mazurskie	0,35	13	0,32	13	0,34	13	0,27	15	0,26	15	0,27	15
Wielkopolskie	0,69	1	0,66	3	0,67	2	0,64	3	0,63	3	0,64	3
Zachodniopomorskie	0,37	11	0,34	11	0,38	11	0,35	12	0,32	12	0,35	11

Źródło: obliczenia własne

Na podstawie analizy rankingów uzyskanych metodą bez wzorca BZW zauważono, że wybór procedury normalizującej zmienne diagnostyczne nie miał znaczącego wpływu na klasyfikacje województw, co potwierdziły wartości współczynnika korelacji rang Spearmana zaprezentowane w tabelach 5 i 6. Uzyskane wyniki uporządkowania województw, dla trzech algorytmów normalizacji, zarówno w 2010, jak i w 2014 roku wskazują na bardzo silną zbieżność.

Tabela 5. Wartości współczynników korelacji rang Spearmana (2010)<sup>3</sup>

BZW 2010	SK	SW	U
SK	1		
SW	0,982	1	
U	0,991	0,976	1

Źródło: obliczenia własne

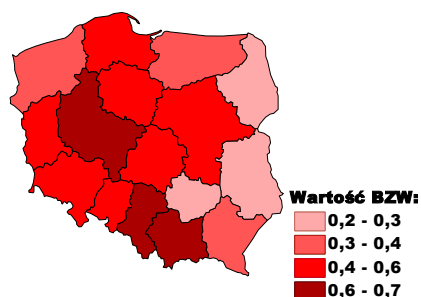
Tabela 6. Wartości współczynników korelacji rang Spearmana (2014)<sup>4</sup>

BZW 2014	SK	SW	U
SK	1		
SW	0,991	1	
U	0,997	0,985	1

Źródło: obliczenia własne

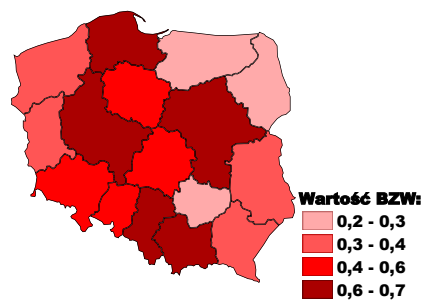
Przestrzenną ilustrację rozmieszczenia wartości miernika BZW dla obu badanych okresów pokazano na mapach Polski (patrz rysunek 3 i 4). Analogicznie do prezentacji SMR, wykorzystano klasyfikację BZW otrzymane z użyciem procedury unitaryzacji, a ciemniejsze kolory województw wskazują na wyższe wartości miernika syntetycznego, czyli wyższy poziom rozwoju społeczno-gospodarczego.

Rysunek 3. Przestrzenne rozmieszczenie wartości miernika BZW w 2010 r. (unitaryzacja)



Źródło: opracowanie własne

Rysunek 4. Przestrzenne rozmieszczenie wartości miernika BZW w 2010 r. (unitaryzacja)



Źródło: opracowanie własne

## PORÓWNANIE WYNIKÓW SMR I BZW

Ostatnim etapem analiz było porównanie uzyskanych wyników otrzymanych odmiennymi algorytmami porządkowania obiektów z różnymi metodami normalizacji cech diagnostycznych.

W tabelach 7 i 8 zestawiono wartości współczynnika korelacji rang Spearmana wyznaczone dla rankingów SMR i BZW dla lat 2010 i 2014.

<sup>3, 32</sup> SK – standaryzacja klasyczna; SW – standaryzacja Webera; U – unitaryzacja.

Tabela 7. Zbieżność rankingów SMR i BZW.

Wyniki współczynnika korelacji rang dla 2010 roku

Rok 2010		BZW		
		SK	SW	U
SMR	SK	0,982	0,959	0,974
	SW	0,897	0,909	0,882
	U	0,982	0,953	0,979

Źródło: obliczenia własne

Tabela 8. Zbieżność rankingów SMR i BZW.

Wyniki współczynnika korelacji rang dla 2014 roku

Rok 2010		BZW		
		SK	SW	U
SMR	SK	0,956	0,959	0,959
	SW	0,941	0,959	0,935
	U	0,938	0,935	0,941

Źródło: obliczenia własne

Uzyskane wartości współczynników korelacji świadczą o tym, iż najwyższym stopniem skorelowania w 2010 wyróżniały się rankingi SMR i BZW konstruowane w oparciu o standaryzację klasyczną i unitaryzację (0,974-0,982). Nieco niższe rezultaty na poziomie 0,882-0,959 otrzymano dla klasyfikacji SMR i BZW budowanych przy użyciu standaryzacji Webera. Można to tłumaczyć wystąpieniem w rozkładach zmiennych  $X_2$  i  $X_6$  wartości odstających oraz silnej asymetrii, co wpłynęło na końcową hierarchię województw. W 2014 roku zbieżność klasyfikacji regionów zbudowanych miernikami SMR i BZW była bardzo silna w przypadku wszystkich użytych formuł normalizacji zmiennych.

W roku 2010 na pierwszych miejscach uplasowały się województwa: Małopolskie, Śląskie, Wielkopolskie oraz Mazowieckie. Najniższy poziom rozwoju społeczno-gospodarczego zdiagnozowano dla województwa Kujawsko-Pomorskiego. Po czterech latach obserwacji zauważono niewielkie różnice w klasyfikacji regionów. W czołówce nadal pozostały te same województwa, jednak znacząco spadła pozycja Podlasia.

## PODSUMOWANIE

Z przeprowadzonego badania wynika, iż korzystanie z różnych metod normalizacji cech diagnostycznych może skutkować otrzymaniem alternatywnej klasyfikacji województw, która nie wynika ze zmiany struktury danych. Modyfikacja algorytmów rankingów na etapie doprowadzania zmiennych do ich wzajemnej porównywalności może przyczyniać się do różnic w uporządkowaniu województw w danym roku, co nie jest powiązane ze zmianą poziomu rozwoju regionu. Ponadto subiektywizm w konstrukcji taksonomicznych mierników



syntetycznych rzutuje na hierarchię województw w rankingach. W literaturze można znaleźć wiele wskazówek na temat metod doboru cech, czy procedur budowy klasyfikacji regionów od których zależą wyniki badań oraz poprawność wniosków. Uwzględnienie opinii ekspertów wraz ze szczegółową analizą zjawiska wielo cechowego jakim jest rozwój społeczno-gospodarczy może gwarantować prawidłowe uporządkowanie badanych obiektów.

## BIBLIOGRAFIA

- Bartosiewicz S. (1976) Propozycja metody tworzenia zmiennych syntetycznych. *Prace Akademii Ekonomicznej we Wrocławiu*, 84, 5-7.
- Cieślak M. (1974) Taksonomiczna procedura prognozowania rozwoju gospodarczego i określenia potrzeb na kadry kwalifikowane. *Przegląd Statystyczny*, 21(1), 29-39.
- Chrzanowska M., Drejerska N. (2016) Ocena poziomu rozwoju społeczno - gospodarczego gmin województwa mazowieckiego z wykorzystaniem metod analizy wielowymiarowej. *Wiadomości Statystyczne*, 6, 59-69.
- Gatnar E., Walesiak M. (red.) (2004) *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*. Wydawnictwo Akademii Ekonomicznej im. O. Langego, Wrocław.
- Hellwig Z. (1968) Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom rozwoju oraz zasoby i strukturę wykwalifikowanych kadr. *Przegląd Statystyczny*, 15(4), 307-326.
- Kisieleńska J. (2017) Ranking województw ze względu na potencjał rolnictwa. *Roczniki Naukowe Ekonomii Rolnictwa i Rozwoju Obszarów Wiejskich*, 104(1), 56-71.
- Kudelko J. (2013) Uwarunkowania i kierunki rozwoju województw Polski Wschodniej jako regionów słabo rozwiniętych. *Studia Komitetu Przestrzennego Zagospodarowania Kraju PAN*, 151, Warszawa, KPZK PAN.
- Kukuła K. (1986) Propozycja miary zgodności układów porządkowych. *Zeszyty Naukowe AE, Kraków*.
- Kukuła K. (2000) *Metoda unitaryzacji zerowanej*, PWN, Warszawa.
- Kukuła K. (2012) Propozycja budowy rankingu obiektów z wykorzystaniem cech ilościowych oraz jakościowych [w:] *Metody Ilościowe w Badaniach Ekonomicznych*, XIII/1, 5-16.
- Kukuła K. (2014) Budowa rankingu województw ze względu na wyposażenie techniczne rolnictwa w Polsce. *Wiadomości Statystyczne*, 7, 62-76.
- Kukuła K., Luty L. (2015) Propozycja procedury wspomagającej wybór metody porządkowania liniowego. *Przegląd Statystyczny*, 62 (2), 219-231.
- Kurkiewicz J., Pocięcha J., Zając K. (1991) *Metody wielowymiarowej analizy porównawczej w badaniach rozwoju demograficznego*. Szkoła Główna Handlowa, Instytut Statystyki i Demografii, Monografie i Opracowania, 336, Warszawa.
- Malina A., Zeliaś A. (1997) O budowie taksonomicznej miary jakości życia. Syntetyczna miara rozwoju jest narzędziem statystycznej analizy porównawczej. *Taksonomia*, 4, 238-262.

- Mazur A., Witkowska D. (2006) Zastosowanie wybranych mierników taksonomicznych do oceny nieruchomości. *Zeszyty Naukowe SGGW–Ekonomika i Org. Gosp. Żywnościowej*, 60, 251-258.
- Sen A. K. (2002) *Rozwój i wolność. Zysk i s-ka*, Poznań.
- Stec M. (2011) Uwarunkowania rozwojowe województw w Polsce – analiza statystyczno-ekonometryczna. *Nierówności Społeczne a Wzrost Gospodarczy*, 20, 232-251.
- Strahl D. (1978) Propozycja konstrukcji miary syntetycznej. *Przegląd Statystyczny*, 25 (2), 205-215.
- Strahl D. (red.) (2006) *Metody oceny rozwoju regionalnego*. Wydawnictwo Akademii Ekonomicznej we Wrocławiu.
- Szewczuk A. (2011) *Rozwój lokalny i regionalny-główne determinanty*. [w:] Szewczuk A., Kogut Jaworska M., Ziolo M., *Rozwój lokalny i regionalny. Teoria i praktyka*. Wydawnictwo CH Beck, Warszawa.
- Szymła Z. (2000) *Determinanty rozwoju regionalnego*. Wydawnictwo Zakład Narodowy im. Ossolińskich, Wrocław–Warszawa–Kraków.
- Szymła Z. (2005) Podstawy badań rozwoju regionalnego. *Zeszyty Naukowe Wyższej Szkoły Ekonomicznej w Bochni*, 3, 101-111.
- Walesiak M. (2003) Uogólniona miara odległości GDM jako syntetyczny miernik rozwoju w metodach porządkowania liniowego. *Prace Naukowe Akademii Ekonomicznej we Wrocławiu, Taksonomia*, 10, 988, 134-144.
- Walesiak M. (2014) Przegląd formuł normalizacji wartości zmiennych oraz ich własności w statystycznej analizie wielowymiarowej. *Przegląd Statystyczny*, 61, 4, 363-372.
- Zielińska-Sitkiewicz M. (2017) Classification of Selected Food Industry Companies Listed on the Warsaw Stock Exchange - the Impact of Normalization Procedures. *Economic Science for Rural Development Conference Proceedings*, 46, 359-366.

**CLASSIFICATION METHODS IN COMPARATIVE ANALYSIS  
OF THE SOCIO-ECONOMIC DEVELOPMENT OF POLISH  
VOIVODESHIPS IN 2010 AND 2014 – THE IMPACT OF DATA  
NORMALIZATION PROCEDURES ON RESULTS OF RANKING**

**Abstract:** The main aim of the paper is analysis the impact of changes of the classification methods and formulas of the data normalization on the result of the linear ordering. In the present study uses synthetic taxonomic measure SMR and the relative level of development BZW to compare the socio-economic development level of Polish voivodeships in 2010 and 2014. In the process of multidimensional comparative analysis the formulas of the classical standardization, Weber standardization and unitarisation were used.

**Keywords:** ranking, data normalization, synthetic taxonomic measures, SMR, BZW, socio-economic development

## METODY ANALIZY TRWANIA W IDENTYFIKACJI DETERMINANT RYZYKA WYKREŚLENIA Z REJESTRU BEZROBOTNYCH

**Iwona Markowicz (ORCID: 0000-0003-1119-0789)**  
**Beata Bieszk-Stolorz (ORCID: 0000-0001-8086-9037)**  
Wydział Nauk Ekonomicznych i Zarządzania  
Uniwersytet Szczeciński  
e-mail: iwona.markowicz@wneiz.pl, beatus@wneiz.pl

**Streszczenie:** Rejestr bezrobotnych może stanowić bazę danych do badania sytuacji na regionalnym rynku pracy. Ważne są w tym przypadku zarówno przyczyna bezrobocia, jak i powód wyrejestrowania. Osoby, które nie zgłosiły się w urzędzie w wyznaczonym terminie lub nie przyjęły propozycji pracy stanowią dużą grupę wśród zarejestrowanych. Stąd podjęto próbę identyfikacji cech wpływających na ryzyko wykreślenia z rejestru. Celem artykułu jest ocena ryzyka wykreślenia z rejestru i identyfikacja jego determinant. Analizie poddano czas pozostawania w rejestrze bezrobotnych, stąd wykorzystano metody analizy trwania.

**Słowa kluczowe:** model regresji Coxa, tablice trwania, funkcja hazardu, bezrobocie

### WSTĘP

Wśród osób rejestrujących się w Powiatowych Urzędach Pracy są osoby rzeczywiście poszukujące tej pracy i potrzebujące pomocy w jej znalezieniu, ale są i takie osoby, które z tej pomocy korzystać nie chcą. Zarejestrowani, którzy nie wyrażą zgody na określenie profilu pomocy, odmówią bez uzasadnionej przyczyny przyjęcia propozycji odpowiedniej pracy lub innej formy pomocy ze strony PUP, poddania się badaniom lekarskim lub psychologicznym, mającym na celu ustalenie zdolności do pracy, odmówią bez uzasadnionej przyczyny udziału w działaniach w ramach Programu Aktywizacja i Integracja, realizowanych przez PUP, nie stawiają się w urzędzie w wyznaczonym terminie, z własnej winy przerwą staż, szkolenie, wykonywanie prac społecznie użytecznych, realizację indywidualnego planu działania lub inną formę

pomocy, z własnej winy przerwą program przygotowania zawodowego dorosłych są z rejestru wykreślane [Ustawa z dnia 20 kwietnia 2004] bez podania przyczyny przez osobę zainteresowaną (za pierwszym razem jest to wykreślane na 120 dni). Wykreśleni stanowią dużą grupę wśród wyrejestrowanych bezrobotnych. Stąd podjęto próbę identyfikacji cech wpływających na ryzyko wykreślenia z rejestru.

Celem artykułu jest ocena ryzyka wykreślenia i identyfikacja jego determinant. Źródłem danych był rejestr bezrobotnych Powiatowego Urzędu Pracy w Szczecinie. Wśród analizowanych cech wzięto pod uwagę płeć, wiek, wykształcenie, staż pracy i liczbę kolejnych zarejestrowań osoby bezrobotnej. Przeprowadzone badanie pozwoliło na określenie grup osób bezrobotnych bardziej podatnych na rezygnację z pośrednictwa urzędu bez podania przyczyny.

## WYKORZYSTANE DANE

Dane wykorzystane w badaniu, którego wyniki są prezentowane w niniejszym artykule, pochodzą z rejestru Powiatowego Urzędu Pracy w Szczecinie. Obejmują one dane indywidualne 22078 osób rejestrujących się jako bezrobotne w 2013 roku. Obserwacja trwała do końca 2014 roku. Informacje o danych liczbowych zawarto w tabeli 1.

Tabela 1. Liczba zarejestrowanych bezrobotnych w PUP w Szczecinie w 2013 roku i obserwowanych do końca 2014 roku według cech (zmiennych w modelach)

Zmienna	Wariant	Symbol	Liczba wykreślonych (1)	Liczba pozostałych (0)	Suma
Ogółem			8965	13113	22078
Płeć	mężczyzna	M	5701	6607	12308
	kobieta	K	3264	6506	9770
Wiek	18-24 lata	W1	2257	1891	4148
	25-34 lata	W2	2966	4390	7356
	35-44 lata	W3	1734	2525	4259
	45-54 lata	W4	1214	2283	3497
	55-60 lat	W5	629	1556	2185
	60-65 lat	W6	165	468	633
Wykształcenie	co najwyżej gimnazjalne	S1	2932	2191	5123
	zasadnicze zawodowe	S2	2220	2796	5016
	średnie ogólnokształcące	S3	1223	1636	2859
	średnie zawodowe	S4	1415	2671	4086
	wyższe	S5	1175	3819	4994
Zarejestrowanie	pierwszy raz	Z0	2045	3373	5418
	kolejny raz	Z1	6920	9740	16660
Staż pracy	brak stażu	D0	4195	3362	7557
	posiadany staż	D1	4770	9751	14521

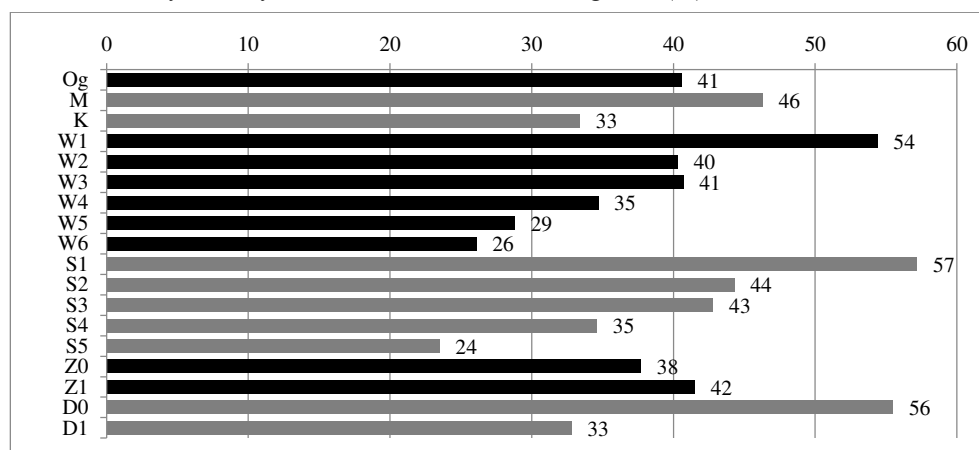
Źródło: opracowanie własne

Poza ogólną liczbą badanych, podane zostały informacje dotyczące cech osób bezrobotnych, takich jak: płeć, wiek, wykształcenie, kolejność zarejestrowania i staż pracy. Cechy te stanowią zmienne w szacowanych modelach, dlatego też wskazano zarówno warianty cech jak i ich symbole. Ostatnia kolumna

wskazuje liczbę osób bezrobotnych rejestrujących się w 2013 roku. Na tą sumę składają się liczba wykreślonych do końca 2014 roku oraz liczba pozostałych osób. Ponieważ przedmiotem zainteresowania w niniejszym badaniu jest wykreślenie bezrobotnego z rejestru, to stanowi ono obserwację pełną i zostało oznaczone wartością 1. Pozostałe przypadki stanowią tu obserwacje cenzurowane i przyjęto dla nich wartość 0 [Hosmer, Lemeshow 1999]. Wśród badanych bezrobotnych przeważają mężczyźni (56%), osoby w wieku 25-34 lata (33%), osoby z najniższym wykształceniem, czyli co najwyżej gimnazjalnym (23%), rejestrujące się po raz kolejny (75%) oraz posiadające już doświadczenia zawodowe (66%).

Udział bezrobotnych wykreślonych z rejestru jest stosunkowo wysoki, wynosi 41% wszystkich rejestrujących się w 2013 roku w PUP w Szczecinie. Udział ten według poszczególnych wariantów analizowanych zmiennych zaprezentowano na rysunku 1. Wykreślenie, jako sposób wyjścia z rejestru przeważał (ponad 50%) w następujących wariantach cech: S1 – wykształcenie co najwyżej gimnazjalne, D0 – brak stażu pracy, W1 – wiek 18-24 lata.

Rysunek 1. Udział zarejestrowanych bezrobotnych w PUP w Szczecinie w 2013 roku i wykreślonych do końca 2014 roku według cech (%)



Źródło: opracowanie własne

## METODA BADAWCZA

Analizie poddano czas pozostawania w rejestrze osób bezrobotnych, stąd wykorzystano wybrane metody analizy trwania: model regresji (hazardu) Coxa i funkcję hazardu jako jeden z elementów tablicy trwania. Metody te można stosować w przypadku występowania danych cenzurowanych. Tak właśnie jest w omawianym badaniu. Wszystkie osoby bezrobotne zarejestrowane w 2013 roku i niewykreślone do końca 2014 roku stanowią właśnie dane cenzurowane (zakodowane jako 0).

Rozkład czasu pozostawania w rejestrze bezrobotnych może być wyrażony za pomocą funkcji hazardu  $h(t)$ . Jest ona stosunkiem funkcji gęstości prawdopodobieństwa  $f(t)$  i funkcji przetrwania  $S(t)$  [Balicki 2006]. Funkcja hazardu określa ryzyko wystąpienia określonego zdarzenia w krótkim przedziale czasu  $(t; t + \Delta t)$ , pod warunkiem, że nie wystąpiło ono do czasu  $t$  [Bieszk-Stolorz, Markowicz 2012].

Omawiane metody służą właśnie do wyznaczenia funkcji hazardu. Analizy intensywności wykreśleń bezrobotnych według ich cech dokonano przy zastosowaniu modelu Coxa [Cox, Oakes 1984], nazywanego modelem proporcjonalnego hazardu [szerzej na temat zastosowania modelu czasu trwania w naukach ekonomicznych: Bieszk-Stolorz 2013, Markowicz 2012, Landmesser 2013, Frątczak, Gach-Ciepiela, Babiker 2005, Gruszczyński 2012, Ostasiewicz 2000]. Model ten można zapisać następująco:

$$h(t; x_1, x_2, \dots, x_n) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n) \quad (1)$$

gdzie:  $x_1, x_2, \dots, x_n$  – zmienne niezależne,  $h_0(t)$  – hazard odniesienia lub zerowa linia hazardu,  $\beta_1, \beta_2, \dots, \beta_n$  – współczynniki modelu,  $t$  – czas obserwacji.

Parametry tego modelu są wyznaczone przez maksymalizację funkcji częściowej wiarygodności [Cox 1972, Cox 1975, Cox, Oakes 1984]. Jest to model semiparametryczny, ponieważ funkcja hazardu jest iloczynem niewyspecyfikowanej parametrycznie bazowej funkcji hazardu i funkcji zmiennych objaśniających, dla których szacowane są parametry.

Zdarzeniem początkowym była rejestracja osoby bezrobotnej w 2013 roku, a zdarzenie końcowe to wykreślenie tej osoby z rejestru. Jeżeli wykreślenie nie nastąpiło, to taką obserwację uznano za cenzurowaną.

Zastosowany sposób definiowania (kodowania) 0-1 zmiennych (dla przekształconych zmiennych dychotomicznych), pozwala na wyznaczenie  $n - 1$  parametrów dla  $n$  zmiennych (uniknięcie współliniowości) [Bieszk-Stolorz, Markowicz 2012]. Jedną, wybraną zmienną staje się punktem odniesienia. W przeprowadzonej analizie jako grupę referencyjną w przypadku poszczególnych cech przyjęto: mężczyzn, bezrobotnych w wieku od 18 do 24 lata, bezrobotnych z wykształceniem co najwyżej gimnazjalnym, zarejestrowanych po raz pierwszy i nie posiadających doświadczenia zawodowego. Wartość  $\exp(\beta)$  jest zatem interpretowana jako ryzyko względne (*hazard ratio* – *HR*), czyli ryzyko zajścia zdarzenia dla danych dwóch jednostek. Jeśli jedna z nich należy do określonej grupy ( $j$ ), a druga do grupy referencyjnej (0), to iloraz hazardu można zapisać wzorem:

$$HR_{j/0} = \frac{h_0 \exp\left(\sum_{i=1}^n \beta_i x_{ij}\right)}{h_0 \exp\left(\sum_{i=1}^n \beta_i x_{i0}\right)} = \exp(\beta_j) \quad (2)$$

Funkcję hazardu jako jeden z elementów tablicy trwania, można zapisać wzorem [Markowicz 2012]:

$$\hat{h}_t = \frac{w_t}{\left(n_t^* - \frac{w_t}{2}\right)a_t} \quad (3)$$

gdzie:  $t$  – początek przedziału czasu trwania  $\langle t, t+1 \rangle$ ,  $n_t^*$  – liczba bezrobotnych narażonych w przedziale,  $w_t$  – liczba bezrobotnych wykreślonych w przedziale,  $a_t$  – długość przedziału czasu trwania (przyjęto miesięczne przedziały).

## WYNIKI BADAŃ

W celu identyfikacji determinant ryzyka wykreślenia z rejestru bezrobotnych zastosowano funkcje hazardu: w modelu Coxa (wskazanie ryzyka względnego) oraz w tablicach trwania (miesięczna stopa hazardu). W tabeli 2 przedstawiono wyniki estymacji modelu Coxa dla badanych bezrobotnych ogółem ( $\chi^2 = 2241,19$ ;  $p = 0,0000$ ;  $p(\beta_i) < 0,05$ ). Elementami wykorzystywanymi w niniejszej analizie są hazard względny oraz jego przedział ufności. Zostały one przedstawione na rysunku 2. Dla wszystkich zmiennych pierwszy wariant stanowi grupę referencyjną (tabela 1). Zatem należy stwierdzić, że ryzyko wykreślenia z rejestru dla kobiet jest mniejsze o 34% niż dla mężczyzn (wartość hazardu poniżej 1). Uwzględniając wiek i wykształcenie, grupą o największym ryzyku są bezrobotni najmłodsi oraz najslabiej wykształceni. Częściej wykreślani są bezrobotni rejestrujący się po raz kolejny oraz mający doświadczenie zawodowe (posiadający staż pracy).

Model regresji Coxa wyznaczono także dla kobiet ( $\chi^2 = 604,80$ ;  $p = 0,0000$ ;  $p(\beta_i) < 0,05$ ;  $p(S2) > 0,05$ ) i mężczyzn ( $\chi^2 = 1246,46$ ;  $p = 0,0000$ ;  $p(\beta_i) < 0,05$ ). Wyniki przedstawiono na rysunkach 3 i 4. Wnioski są w zasadzie podobne jak w analizie ogółem, z małymi wyjątkami. Parametr przy zmiennej S2 dla kobiet jest nieistotny, co oznacza, że ryzyko wykreślenia dla kobiet z wykształceniem zasadniczym zawodowym nie różni się istotnie od ryzyka dla kobiet z wykształceniem co najwyżej gimnazjalnym. Mężczyźni w wieku 60-64 lata (brak kobiet w tej grupie wieku) stanowią drugą grupę ryzyka wykreślenia zaraz po najmłodszych mężczyznach.

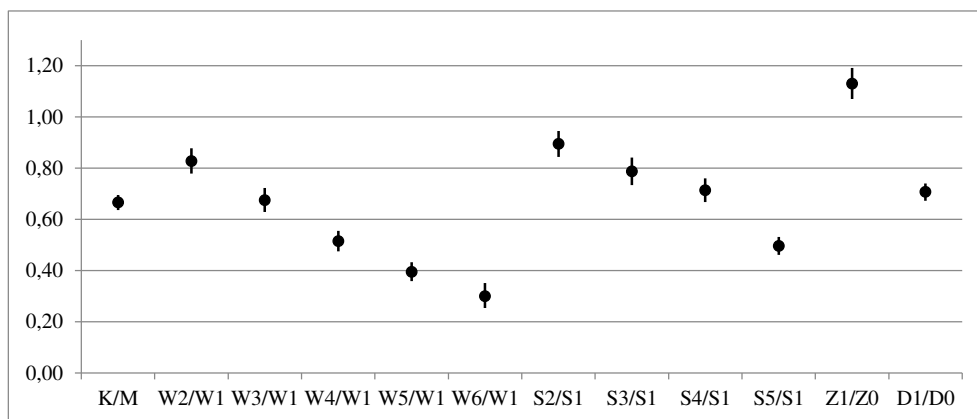
Tabela 2. Wyniki estymacji modelu (1) dla bezrobotnych ogółem

Zmienna	Ocena parametru	Błąd standardowy	Statystyka Walda	Poziom $p$	Hazard względny	Hazard względny przedział ufności 95%	
K/M	-0,4083	0,0226	326,7747	0,0000	0,6648	0,6360	0,6949
Z1/Z0	0,1216	0,0273	19,8263	0,0000	1,1294	1,0705	1,1915
W2/W1	-0,1903	0,0306	38,5943	0,0000	0,8267	0,7785	0,8779
W3/W1	-0,3945	0,0357	122,0670	0,0000	0,6740	0,6284	0,7229
W4/W1	-0,6665	0,0393	288,1717	0,0000	0,5135	0,4755	0,5546
W5/W1	-0,9315	0,0481	375,1093	0,0000	0,3940	0,3585	0,4329
W6/W1	-1,2084	0,0827	213,5862	0,0000	0,2987	0,2540	0,3512

Zmienna	Ocena parametru	Błąd standardowy	Statystyka Walda	Poziom $p$	Hazard względny	Hazard względny przedział ufności 95%	
S2/S1	-0,1126	0,0288	15,2948	0,0001	0,8935	0,8444	0,9454
S3/S1	-0,2408	0,0347	48,1913	0,0000	0,7860	0,7343	0,8413
S4/S1	-0,3394	0,0330	105,6640	0,0000	0,7122	0,6675	0,7598
S5/S1	-0,7029	0,0366	368,1418	0,0000	0,4952	0,4609	0,5320
D1/D0	-0,3481	0,0242	206,9950	0,0000	0,7060	0,6733	0,7403

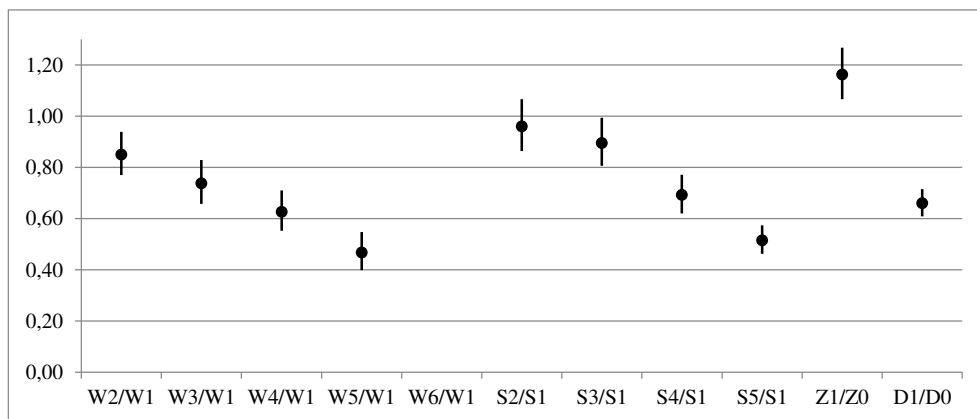
Źródło: opracowanie własne (oznaczenia jak w tabeli 1)

Rysunek 2. Hazard względny wykreślenia z rejestru dla bezrobotnych ogółem



Źródło: opracowanie własne

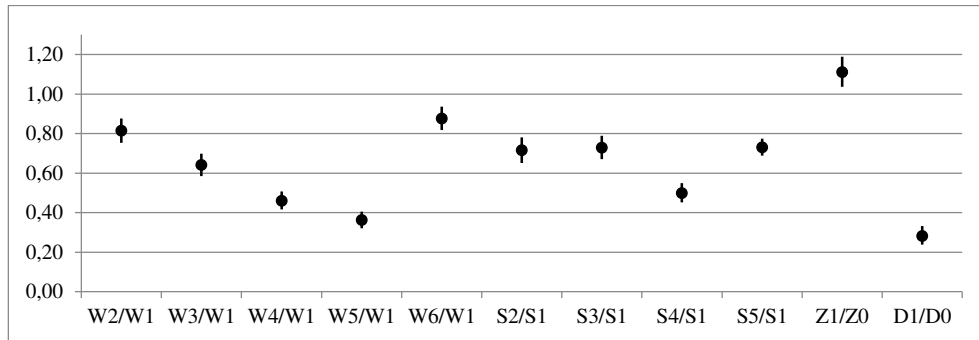
Rysunek 3. Hazard względny wykreślenia z rejestru dla bezrobotnych kobiet



Źródło: opracowanie własne

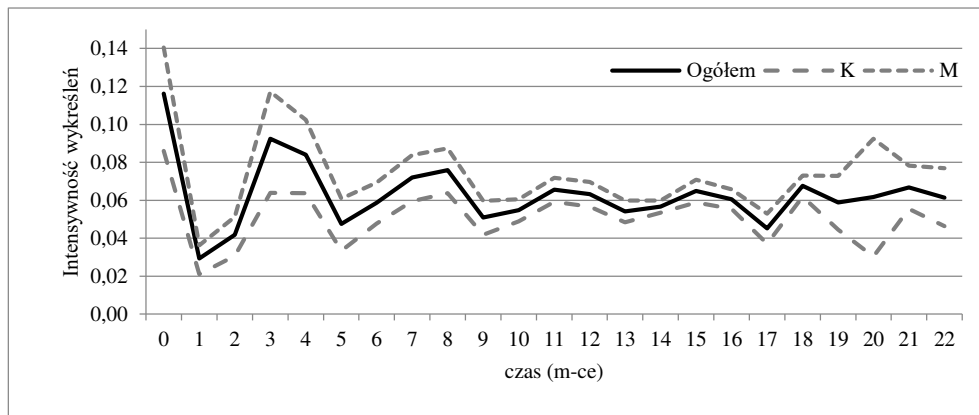


Rysunek 4. Hazard względny wykreślenia z rejestru dla bezrobotnych mężczyzn



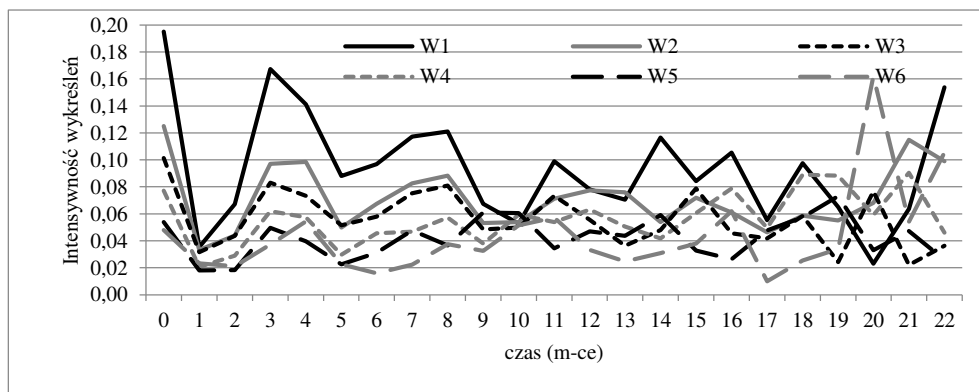
Źródło: opracowanie własne

Rysunek 5. Intensywność wykreśleń z rejestru bezrobotnych ogółem oraz według płci



Źródło: opracowanie własne

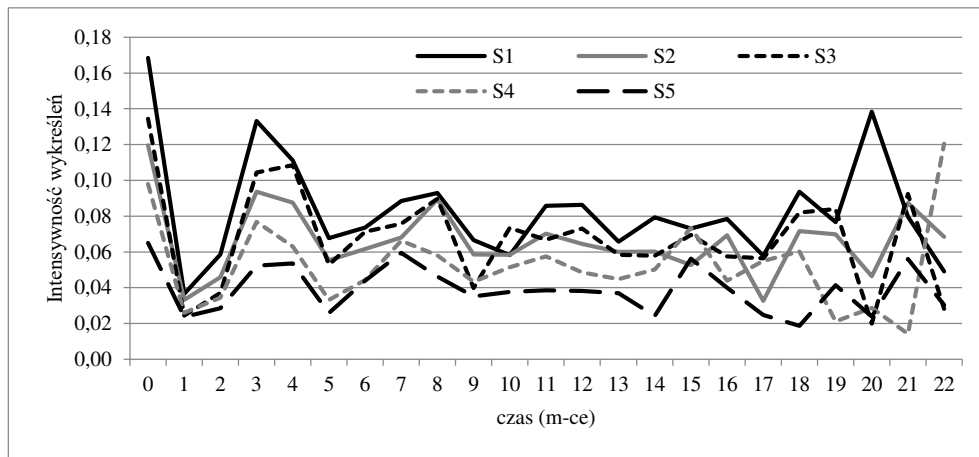
Rysunek 6. Intensywność wykreśleń z rejestru bezrobotnych według wieku



Źródło: opracowanie własne

Drugim etapem badania była budowa tablic trwania w bezrobociu dla bezrobotnych ogółem oraz bezrobotnych według płci, wieku, wykształcenia, kolejności rejestracji w urzędzie pracy oraz posiadania stażu pracy. Utworzono tablice dla miesięcznych przedziałów czasu. Wyznaczona funkcja hazardu pozwala na porównanie ryzyka wykreślenia danej grupy bezrobotnych z inną grupą oraz na obserwację zmian ryzyka wykreślenia wraz z upływem czasu od rejestracji. Wyniki przedstawiono na rysunkach 5-7 (na osi czasu podano początek przedziału czasu trwania bezrobocia). Szczególnie wysoka intensywność wykreśleń miała miejsce w pierwszym (przedział 0-1) i czwartym miesiącu trwania bezrobocia. Jest to początek okresu posiadania statusu bezrobotnego i okres wezwań do urzędu pracy. Niestawianie się na wezwania już na początku trwania bezrobocia sugeruje brak gotowości do podjęcia pracy. Większe ryzyko wykreśleń w całym badanym okresie dotyczyło mężczyzn niż kobiet. Biorąc pod uwagę wiek i wykształcenie bezrobotnych, można stwierdzić, że ryzyko wykreśleń maleje wraz ze wzrostem wieku i wykształcenia. Zatem to osoby młode i słabo wykształcone rejestrują się w urzędzie, ale nie oczekują pomocy w znalezieniu zatrudnienia czy uczestniczenia w szkoleniu, stażu. W przypadku grup wieku szczególnie wysoka intensywność wykreśleń dotyczy osób młodych (18-24 lata). W przypadku wykształcenia należy podkreślić stosunkowo wysoką intensywność osób z wykształceniem ogólnokształcącym (często wyższą niż osób z niższym wykształceniem – zasadniczym zawodowym).

Rysunek 7. Intensywność wykreśleń z rejestru bezrobotnych według wykształcenia



Źródło: opracowanie własne

Funkcje hazardu wyznaczono również dla grup bezrobotnych według kolejności zgłoszenia oraz posiadanego doświadczenia zawodowego. Intensywność wykreśleń dla osób rejestrujących się pierwszy i kolejny raz były dość podobne.

Natomiast intensywność wykreśleń dla osób bez stażu pracy była w całym okresie wyższa niż dla osób posiadających już staż.

## PODSUMOWANIE

Przeprowadzona analiza dotyczyła osób zarejestrowanych jako bezrobotne w 2013 roku w PUP w Szczecinie, ale które nie zgłosiły się w urzędzie w wyznaczonym terminie lub nie przyjęły propozycji pracy i zostały pozbawione statusu osoby bezrobotnej. Stąd podjęto próbę identyfikacji cech wpływających na ryzyko wykreślenia z rejestru (płeć, wiek, wykształcenie, kolejność zarejestrowania i staż pracy). Wyniki badań wskazują na następujące wnioski:

1. bezrobotni wykreśleni z rejestru stanowili 41% wszystkich rejestrujących się,
2. ponad 50% udział wykreśleń dotyczył grup: z wykształceniem co najwyżej gimnazjalnym (57%), rejestrujących się pierwszy raz (56%) i w wieku 18-24 lata (54%),
3. ryzyko wykreślenia z rejestru dla kobiet jest mniejsze o 34% niż dla mężczyzn,
4. grupą według wieku o największym ryzyku są najmłodsi bezrobotni (18-24 lata)
5. grupą według wykształcenia o największym ryzyku są osoby najslabiej wykształcone (co najwyżej gimnazjalne),
6. częściej wykreśleni są bezrobotni rejestrujący się po raz kolejny oraz mający doświadczenie zawodowe (posiadający staż pracy),
7. z analizy funkcji hazardu dla miesięcznych przedziałów czasu trwania bezrobocia (tablice trwania) wynika, że szczególnie wysoka intensywność wykreśleń miała miejsce w pierwszym i trzecim miesiącu trwania bezrobocia; jest to początek okresu posiadania statusu bezrobotnego i okres wezwań do urzędu pracy (brak gotowości do podjęcia pracy),
8. większe ryzyko wykreśleń w całym badanym okresie dotyczyło mężczyzn niż kobiet,
9. ryzyko wykreśleń maleje wraz ze wzrostem wieku i wykształcenia,
10. intensywność wykreśleń dla osób rejestrujących się pierwszy i kolejny raz były podobne
11. intensywność wykreśleń dla osób bez stażu pracy była w całym okresie wyższa niż dla osób posiadających już staż.

Zatem to przede wszystkim mężczyźni, osoby młode, słabo wykształcone i bez doświadczeń zawodowych rejestrują się w urzędzie pracy, ale nie oczekują pomocy w znalezieniu zatrudnienia czy udziału w szkoleniu, stażu.

## BIBLIOGRAFIA

- Balicki A. (2006) Analiza przeżycia i tablice wymieralności. PWE, Warszawa.
- Bieszk-Stolorz B. (2013) Analiza historii zdarzeń w badaniu bezrobocia. Volumina.pl Daniel Krzanowski, Szczecin.
- Bieszk-Stolorz B., Markowicz I. (2012) Modele regresji Coxa w analizie bezrobocia. CeDeWu, Warszawa.
- Blanchard O., Diamond P. (1994) Ranking, Unemployment Duration and Wages. *Review of Economic Studies*, 61, 417-434.
- Cox D. R., Oakes D. (1984) *Analysis of Survival Data*. Chapman and Hall, London.
- Fratczak E., Gach-Ciepiela U., Babiker H. (2005) Analiza historii zdarzeń. Elementy teorii, wybrane przykłady zastosowań. SHG, Warszawa
- Gruszczyński M. (red.) (2012) *Mikroekonometria. Modele i metody analizy danych indywidualnych*. Oficyna Wolters Kluwer S. A., Warszawa.
- Hosmer D. W., Lemeshow S. (1999) *Applied Survival Analysis. Regression Modeling of Time to Event Data*. John Wiley & Sons, Inc., New York.
- Landmesser J. (2013) Wykorzystanie metod analizy czasu trwania do badania aktywności ekonomicznej ludności w Polsce. Wydawnictwo SGGW, Warszawa.
- Markowicz I. (2012) Statystyczna analiza żywotności firm. Wydawnictwo Naukowe Uniwersytetu Szczecińskiego.
- Markowicz I., Stolorz B. (2009) Model proporcjonalnego hazardu Coxa przy różnych sposobach kodowania zmiennych. *Przegląd Statystyczny*, (56) 2, 106-115.
- Ostasiewicz S. red. (2000) *Metody oceny i porządkowania ryzyka w ubezpieczeniach życiowych*. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.
- Ustawa z dnia 20 kwietnia 2004 r. o promocji zatrudnienia i instytucjach rynku pracy (Dz. U. z 2017 r. poz. 1065) – art. 33 ust. 4.

**METHODS OF DURATION ANALYSIS IN IDENTIFICATION  
OF DETERMINANTS OF RISK OF ERASURE FROM THE  
UNEMPLOYMENT REGISTRY**

**Abstract:** The unemployment registry may be the database for analysis of the situation on the regional labour market. In this case there are important both the cause of unemployment and the deregistration reason. The lack of clearly defined cause of erasure from the labour office registry makes analyses difficult. Therefore the authors tried to identify the features of these persons influencing the risk of erasure from the unemployment registry. The aim of the paper is assessment of the risk of erasure from the registry and identification of its determinants. Because the time of staying in the unemployment registry was analysed, selected methods of duration analysis were applied.

**Keywords:** the Cox hazard model, duration tables, hazard function, unemployment

## WSPÓŁODDZIAŁYWANIE WYBRANYCH DOPLAT NA WZROST WYDAJNOŚCI PRACY W POLSKICH GOSPODARSTWACH ROLNYCH

**Aleksandra Pawłowska (ORCID: 0000-0001-8964-3624)**

Zakład Zastosowań Matematyki w Ekonomice Rolnictwa  
Instytut Ekonomiki Rolnictwa i Gospodarki Żywnościowej – PIB  
e-mail: [aleksandra.pawlowska@ierigz.waw.pl](mailto:aleksandra.pawlowska@ierigz.waw.pl)

**Streszczenie:** Wydajność czynnika pracy w rolnictwie stanowi jeden z podstawowych wskaźników oddziaływania przy ocenie skuteczności instrumentów wdrażanych w ramach WPR. Celem pracy jest próba identyfikacji i pomiaru efektów wspólnego oddziaływania dopłat do inwestycji oraz rolnośrodowiskowymi i ONW na przyrost wydajności czynnika pracy w polskich gospodarstwach rolnych. W badaniu wykorzystano dane Polskiego FADN dla lat 2007-2015 dotyczące indywidualnych gospodarstw rolnych. Narzędziem badawczym jest metoda Propensity Score Matching.

**Słowa kluczowe:** gospodarstwa rolne, wydajność pracy, polityka rolna, propensity score matching

### WSTĘP

Wydajność czynnika pracy odgrywa szczególną rolę w procesach wzrostu, bowiem „wzrost jej produktywności jest kluczowym czynnikiem dla postępu i przyjęcia przez producentów nowych technologii” [Sielska i in. 2015, s. 8, za: Patra, Nayak 2012]. Pozytywny wpływ może mieć tutaj polityka, w szczególności rozważane w pracy instrumenty Wspólnej Polityki Rolnej (WPR).

Jak wskazuje Rembisz [2007], wzrost wydajności czynnika pracy kształtowany jest przez przyrost oszczędności, wynikających ze wzrostu wynagrodzenia tegoż czynnika. Zmiana oszczędności wpływa z kolei na przyrost inwestycji, a następnie zwiększenie zaangażowania czynnika kapitału rzeczowego u producentów rolnych, prowadząc do przyrostu technicznego uzbrojenia pracy. W rezultacie daje to podstawę w ujęciu per capita do wzrostu wydajności czynnika

pracy. Dla producenta maksymalizującego swoją funkcję celu, wzrost wydajności czynnika wytwórczego prowadzić powinien do wzrostu wynagrodzenia danego czynnika, a w konsekwencji również dochodów producenta [por. np. Krugman, Wells 2012].

Obok generowanego wzrostu oszczędności, źródło wzrostu wydajności czynnika pracy stanowią również transfery budżetowe, w szczególności płatności realizowane w ramach WPR. Zewnętrzny strumień dopływu pieniędzy wpływa, poprzez zwiększenie dochodów producentów rolnych, na przyrost oszczędności będących podstawą inwestowania [Rembisz 2007]. Przyjmując zatem, że typową sytuacją dla producentów rolnych w Polsce jest [por. np. Rembisz, Sielska 2014]:

$$S_{t-1} < I_t, \quad (1)$$

wskazane zależności zapisać można zatem jako:

$$S_{t-1} + B_t \Rightarrow I_t \uparrow \Rightarrow \frac{K_t}{L_t} \uparrow \Rightarrow \frac{y_t}{L_t} \uparrow, \quad (2)$$

gdzie:  $S$  – oszczędności,  $B$  – transfery,  $I$  – inwestycje,  $K$  – czynnik kapitału,  $L$  – czynnik pracy,  $y$  – produkcja,  $t$  – indeks czasu.

Pozytywne implikacje wzrostu wydajności czynnika pracy znajdują swoje odzwierciedlenie również w konstrukcji instrumentów polityki, w szczególności – w przypadku rolnictwa – w mechanizmach Wspólnej Polityki Rolnej. Działaniami wspierającymi inwestycje w polskich gospodarstwach rolnych jest, przykładowo, poddziałanie 4.1 realizowane w ramach PROW 2014-2020 pt. „Wsparcie na inwestycje w gospodarstwach rolnych”, w szczególności operacja typu „Modernizacja gospodarstw rolnych”. Warunkiem kwalifikowalności otrzymania przez gospodarstwa rolne wsparcia jest utrzymanie wzrostu wartości dodanej brutto o co najmniej 10% przez okres 5 lat od dnia przyznania pomocy. Realizacja tego celu nastąpić może poprzez m.in. wzrost wartości dodanej produktu, poprawę jakości produktu czy zmianę struktury produkcji [PROW 2014-2020]. Skala wydatków, tj. przydzielenie na tę operację ok. 18% ogółu środków przeznaczonych na PROW 2014-2020, oraz zainteresowanie beneficjentów świadczą o istotności podejmowanych działań.

Celem badawczym pracy jest zatem pomiar efektu oddziaływania wybranych płatności na wzrost wydajności czynnika pracy. Przedmiotem zainteresowania jest kwantyfikacja wpływu wsparcia dla inwestycji podejmowanych w gospodarstwach rolnych w kontekście korzystania, dodatkowo, z dopłat o charakterze środowiskowym<sup>1</sup>.

<sup>1</sup> W pracy uwzględniono, zgodnie z podziałem stosowanym w Polskim FADN, dopłaty rolnośrodowiskowe i dopłaty do obszarów o niekorzystnych warunkach gospodarowania (ONW).

## METODA BADAWCZA

### Metoda Propensity Score Matching

W pracy rozważany jest wpływ otrzymywania dopłat – jako egzogenicznego czynnika – na wzrost wydajności czynnika pracy w gospodarstwach rolnych. Jak wskazuje Strawiński [2014], metodą wyznaczającą standard oceny efektu oddziaływania wprowadzonego instrumentu polityki jest eksperyment zrandomizowany. Badane gospodarstwa byłyby wówczas losowo przydzielane do grupy eksperymentalnej lub kontrolnej, a więc to, do której grupy należy jednostka byłoby niezależne od wyniku oddziaływania, czyli przyrostu wydajności czynnika pracy [Heckman, Vytlačil 2007]. Zgodnie z podejściem kontrfaktycznym Neymana-Rubina, jednostce przydzielonej do jednej z dwóch grup można przypisać wartość zmiennej wynikowej, tj. wzrost wydajności czynnika pracy, w obu stanach [Guo, Fraser 2015]. Formalnie, zapisać to można jako:

$$Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i}, \quad (3)$$

gdzie:  $Y_i$  – wzrost wydajności czynnika pracy w  $i$ -tym gospodarstwie,  $D_i$  – zmienna binarna przyjmująca wartość 1, gdy  $i$ -te gospodarstwo otrzymało wsparcie lub wartość 0 w przeciwnym przypadku,  $Y_{0i}$ ,  $Y_{1i}$  – wzrost wydajności czynnika pracy w  $i$ -tym gospodarstwie w przypadku odpowiednio nieotrzymania oraz otrzymania wsparcia.

Efekt oddziaływania dopłat dla  $i$ -tego gospodarstwa można byłoby zatem zdefiniować jako:

$$\tau_i = Y_{1i} - Y_{0i}. \quad (4)$$

W praktyce niemożliwym jest jednak zaobserwowanie wyniku dwóch wykluczających się wzajemnie zdarzeń, tj. otrzymania i nieotrzymania dopłat, co określa się mianem tzw. fundamentalnego problemu wnioskowania przyczynowego. Rozwiązaniem jest uwzględnienie oszacowań przybliżających nieobserwowalne wartości zmiennych (tzw. stanów kontrfaktycznych), zgodnie ze wzorem [Szulc 2012]:

$$\tau_i = \begin{cases} Y_{1i} - \hat{Y}_{0i} & \text{dla } D_i = 1 \\ \hat{Y}_{1i} - Y_{0i} & \text{wpp.} \end{cases} \quad (5)$$

gdzie:  $\hat{Y}_{0i}$ ,  $\hat{Y}_{1i}$  – oszacowanie potencjalnej wartości wzrostu wydajności czynnika pracy w  $i$ -tym gospodarstwie w przypadku odpowiednio otrzymania oraz nieotrzymania wsparcia.

Za stan kontrfaktyczny przyjmuje się wartość zmiennej wynikowej dla jednostki (lub jednostek) z grupy kontrolnej, będącej najbardziej podobną pod względem pewnych obserwowalnych charakterystyk do analizowanej  $i$ -tej

jednostki z grupy eksperymentalnej. Wyznaczenie stanów kontrfaktycznych pozwala na pomiar m.in. przeciętnego efektu oddziaływania wobec jednostek poddanych oddziaływaniu (average treatment effect on treated, ATT):

$$ATT = E(\tau | D = 1) = E(Y_1 | D = 1) - E(Y_0 | D = 1)^2. \quad (6)$$

W odniesieniu do rozważanego w pracy zagadnienia, wartość ATT traktować można jako przeciętny wzrost wydajności czynnika pracy wśród beneficjentów programu, wynikający z uzyskiwania danego wsparcia, w porównaniu do gospodarstw nim nieobjętych.

### **Łączenie za pomocą wektora propensity score**

Do wyznaczenia stanu kontrfaktycznego dla jednostki z grupy eksperymentalnej niezbędnym jest znalezienie „podobnej” jednostki w grupie kontrolnej. Aby zminimalizować obciążenie wynikające z selekcji i zapewnić podobieństwo porównywanych grup, jednostki z grupy eksperymentalnej i kontrolnej łączone są na podstawie wektora określonych, obserwowalnych charakterystyk [Trzeciński 2009]. Jak wskazuje się m.in. w pracy Sielskiej i Pawłowskiej [2016] dobór obserwacji kontrolnych z grupy jednostek niepoddanych oddziaływaniu rozważanego czynnika jest kluczowym aspektem przeprowadzanej analizy wpływu.

U podstaw łączenia obserwacji z obu grup leży jednak problem wielowymiarowości, związany z koniecznością łączenia jednostek na podstawie wartości wielu identycznych (lub zbliżonych) cech o często ciągłym charakterze. Jak zauważa Trzeciński [2009] posiadanie bogatego zbioru zmiennych „odpowiedzialnych” za uczestnictwo jednostek w programie jest wymogiem krytycznym dla spełnienia podstawowego założenia metody kontrfaktycznej, tj. warunkowej niezależności wyniku udziału i braku udziału w programie od przynależności jednostki do grupy eksperymentalnej lub kontrolnej<sup>3</sup>.

Sprowadzenie problemu łączenia danych do jednowymiarowego zaproponowali Rosenbaum i Rubin [1983], wprowadzając pojęcie tzw. wektora bilansującego, a jako najprostszą postać funkcji bilansującej wskazali propensity score, zdefiniowaną jako wartość prawdopodobieństwa poddania jednostki oddziaływaniu badanego czynnika [Strawiński 2014]. Metodami wyznaczania propensity score mogą być modele logitowe lub probitowe czy analiza dyskryminacyjna.

---

<sup>2</sup> Przy założeniu, że selekcja, jeżeli występuje, zależy jedynie od obserwowalnych cech jednostek [Strawiński 2014].

<sup>3</sup> Więcej o założeniach analizy kontrfaktycznej piszą m.in. Strawiński [2014] i Szulc [2012].



## WYKORZYSTANE DANE

Badanie przeprowadzono wykorzystując dane Polskiego FADN (Farm Accountancy Data Network) z lat 2007-2015 dla ponad 5 tys. indywidualnych gospodarstw rolnych o różnych typach specjalizacji. Dokonano kwantyfikacji efektu oddziaływania dopłat do inwestycji, rolnośrodowiskowych i ONW na roczny przyrost wydajności czynnika pracy, zdefiniowanej jako wielkość produkcji przypadającej na roczną jednostkę pracy (AWU)<sup>4</sup>. Przyjęto założenie o występowaniu rocznego opóźnienia efektu oddziaływania badanych dopłat na wzrost wydajności czynnika pracy, stąd wybrane obserwowalne charakterystyki gospodarstwa z roku  $t$  wpływały na prawdopodobieństwo otrzymania dopłat w roku  $t+1$ , czego efektem była zmiana wydajności czynnika pracy w roku  $t+2$ .

Wektor propensity score skonstruowano na podstawie modeli logitowych, za pomocą których szacowano wpływ wybranych charakterystyk gospodarstw na prawdopodobieństwo otrzymania analizowanych dopłat. Do modelu jako zmienne objaśniające wybrano taką kombinację ze zbioru 13 zmiennych<sup>5</sup>, dla której współczynnik trafności klasyfikacji był najwyższy [Strawiński 2014, za: Heckman i in. 1997]. Za nadrzędne przyjęto jednakże uzyskanie zbalansowania cech gospodarstw, tak aby zapewnić ich zbliżony rozkład w grupie eksperymentalnej i kontrolnej [Trzeciński 2009]. Stąd też, w przypadku braku możliwości osiągnięcia zbalansowania dla modelu o najwyższej trafności klasyfikacji, do dalszej analizy wybrano model o niższym współczynniku trafności, ale zapewniający zbalansowanie cech między grupami<sup>6</sup>.

Dwie grupy eksperymentalne w badaniu stanowiły gospodarstwa, które w okresie 2008-2014 otrzymały odpowiednio wsparcie dla inwestycji oraz wsparcie dla inwestycji, rolnośrodowiskowe i ONW. Do grupy kontrolnej należały z kolei gospodarstwa, które w całym okresie 2007-2015 nie były beneficjentami analizowanych dopłat.

---

<sup>4</sup> Z uwagi na przedmiot badania, wydajność czynnika pracy nie mogła zostać zdefiniowana jako wartość dodana brutto (GVA) przypadająca na roczną jednostkę pracy (AWU), ponieważ, zgodnie z metodyką FADN, dopłaty do działalności operacyjnej (w tym również dopłaty rolnośrodowiskowe i ONW) są składową wartości dodanej brutto [Floriańczyk i in. 2016].

<sup>5</sup> Do uwzględnionych w modelach logitowych charakterystyk gospodarstw należały: typ specjalizacji, klasa wielkości ekonomicznej, wiek i wykształcenie rolnika, powierzchnia użytków rolnych, średni stan zwierząt, zużycie wewnętrzne w ramach działalności operacyjnej, koszty czynników zewnętrznych, aktywa trwałe i obrotowe, wartość zobowiązań krótko- i długookresowych, średnia wartość kapitału gospodarstwa, inwestycje netto oraz przepływ pieniężny I [patrz Floriańczyk i in. 2016].

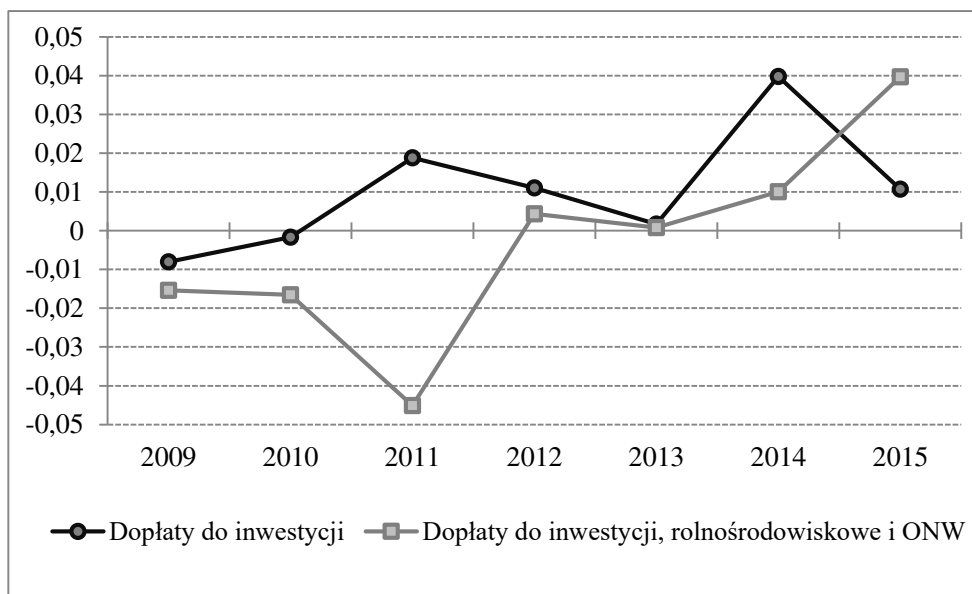
<sup>6</sup> Do specyfikacji wektora propensity score wykorzystano zmienne, dla których modele dla poszczególnych okresów cechowały się trafnością klasyfikacji na poziomie 63-82%.

Wykorzystując metodę Propensity Score Matching zastosowano łączenie 1:1 ze zwracaniem, zatem jedno gospodarstwo z grupy kontrolnej mogło zostać połączone z co najmniej jednym gospodarstwem z grupy eksperymentalnej [Sekhon 2011].

## WYNIKI BADAŃ

Wykorzystując metodę Propensity Score Matching dokonano kwantyfikacji efektu oddziaływania dopłat do inwestycji oraz dopłat do inwestycji, rolnośrodowiskowych i ONW na przyrost wydajności czynnika pracy. Na rysunku 1 pokazano kształtowanie się ATT w latach 2009-2015 w obu przypadkach. Gospodarstwa rolne z grupy eksperymentalnej łączone były z gospodarstwami z grupy kontrolnej na podstawie wartości wektora propensity score, przy zapewnieniu zbalansowania obu grup.

Rysunek 1. Przeciętny efekt oddziaływania (ATT) dopłat do inwestycji oraz dopłat do inwestycji, rolnośrodowiskowych i ONW na przyrost wydajności czynnika pracy w gospodarstwach rolnych w latach 2009-2015



Źródło: opracowanie własne

Uzyskane wyniki wskazują, że wpływ dopłat do inwestycji oraz jednoczesny wpływ dopłat do inwestycji, rolnośrodowiskowych i ONW na wzrost wydajności czynnika pracy wśród beneficjentów programu był dodatni od odpowiednio 2011 i 2012 roku. W przypadku wsparcia dla inwestycji, najwyższa co do wartości

bezwzględnej różnica między gospodarstwami z grupy eksperymentalnej i kontrolnej wystąpiła w 2014 roku. Beneficjenci dopłat otrzymanych w 2013 roku charakteryzowali się o ok. 4 p. proc. wyższym przyrostem wydajności czynnika pracy w 2014 roku niż gospodarstwa należące do grupy kontrolnej. Z kolei najwyższa bezwzględna rozbieżność między gospodarstwami otrzymującymi jednocześnie płatności dla inwestycji, rolnośrodowiskowe i ONW a grupą kontrolną pojawiła się w 2011 roku. Gospodarstwa nieotrzymujące w analizowanym okresie wsparcia, tj. grupa kontrolna, osiągnęły wówczas o ok. 4,5 p. proc. wyższy wzrost wydajności czynnika pracy w porównaniu do gospodarstw, które w poprzednim roku skorzystały z analizowanego wsparcia.

Istotnym dla wyników przeprowadzonego badania jest jednak, że w całym rozważanym okresie (za wyjątkiem 2015 roku) efekt oddziaływania dopłat do inwestycji na zmienną wynikową był wyższy niż efekt jednoczesnego oddziaływania dopłat do inwestycji, rolnośrodowiskowych i ONW. Między analizowanymi trzema rodzajami dopłat nie występuje zatem efekt synergii, który mógłby mieć katalizujący wpływ na przyrost wydajności czynnika pracy w gospodarstwach rolnych. Co więcej, jeżeli łączny wpływ analizowanych dopłat jest niższy niż efekt oddziaływania wyłącznie dopłat do inwestycji, to wysunąć można wniosek o „blokowaniu” wzrostu wydajności czynnika pracy przez korzystanie z dopłat o charakterze środowiskowym.

Uzyskane wyniki, a więc wyższy przyrost wydajności pracy w gospodarstwach, które można sklasyfikować jako konwencjonalne, w porównaniu do gospodarstw ekologicznych są zgodne z wnioskami z prac m.in. Komorowskiej [2012] i Łuczki [2016].

## PODSUMOWANIE

Podstawę rozważań w niniejszej pracy stanowiła wydajność czynnika pracy w polskich gospodarstwach rolnych, na wzrost której, opierając się na mikropodstawach, katalizujący wpływ winny mieć inwestycje. Celem badania było porównanie wpływu działań WPR ukierunkowanych na wsparcie inwestycji w gospodarstwach rolnych w kontekście gospodarstw korzystających dodatkowo z płatności o charakterze środowiskowym. Wykorzystując metodę Propensity Score Matching dokonano kwantyfikacji efektu oddziaływania dopłat do inwestycji, rolnośrodowiskowych i ONW na przyrost wydajności czynnika pracy. Badanie przeprowadzono dla indywidualnych gospodarstw rolnych uczestniczących w Polskim FADN w latach 2007-2015. Potrzeba ewaluacji wdrożonych instrumentów wynika bowiem m.in. z konieczności pomiaru realnych korzyści każdej polityki w kontekście poniesionych na jej prowadzenie kosztów.

Uzyskane wyniki wskazały, że wpływ dopłat do inwestycji oraz jednoczesny wpływ dopłat do inwestycji, rolnośrodowiskowych i ONW na wzrost wydajności czynnika pracy wśród beneficjentów programu był dodatni od odpowiednio 2011 i 2012 roku. W przypadku wsparcia dla inwestycji, najwyższa co do wartości

bezwzględnej różnica między gospodarstwami z grupy eksperymentalnej i kontrolnej wystąpiła w 2014 roku. Z kolei najwyższa bezwzględna rozbieżność między gospodarstwami otrzymującymi jednocześnie płatności dla inwestycji, rolnośrodowiskowe i ONW a grupą kontrolną pojawiła się w 2011 roku. Istotnym dla wyników przeprowadzonego badania jest jednak, że w całym rozważanym okresie (za wyjątkiem 2015 roku) między analizowanymi trzema rodzajami dopłat nie wystąpił efekt synergii, który mógłby mieć dodatkowy wpływ na przyrost wydajności czynnika pracy w gospodarstwach rolnych. Co więcej, jeżeli łączny wpływ analizowanych dopłat okazał się niższy niż efekt oddziaływania wyłącznie dopłat do inwestycji, to mówić można o swego rodzaju „blokowaniu” wzrostu wydajności czynnika pracy przez korzystanie z dopłat o charakterze środowiskowym.

Ocena ta wydaje się być szczególnie istotna w kontekście instrumentów obecnego PROW oraz wymogu wykazania przez beneficjentów dopłat do inwestycji wzrostu wartości dodanej brutto o co najmniej 10% przez okres 5 lat od dnia przyznania pomocy. Z uwagi na znaczenie kwestii środowiskowych, co znajduje swoje odzwierciedlenie w konstrukcji PROW 2014-2020, wstępną rekomendacją jest zatem uelastyczenie wymogów stosowanych wobec beneficjentów programu, w zależności od charakteru (konwencjonalnego lub ekologicznego) gospodarstwa.

## BIBLIOGRAFIA

- Floriańczyk Z., Osuch D., Płonka R. (2016) Wyniki Standardowe 2015 uzyskane przez gospodarstwa rolne uczestniczące w Polskim FADN. Część I. Wyniki Standardowe. IERiGŻ-PIB, Warszawa.
- Guo S., Fraser M. W. (2015) Propensity Score Analysis. Statistical Methods and Applications. Second Edition, Sage Publications.
- Heckman J., Ichimura H., Todd P. (1997) Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program. *The Review of Economic Studies*, 64(4), 605-654.
- Heckman J., Vytlacil E. (2007) Econometric Evaluation of Social Programs, Part I: Casual Models, Structural Models and Econometric Policy Evaluation. [w:] Heckman J., Leamer E. (red.) *Handbook of Econometrics*. 6B, Elsevier, Amsterdam, 4779-4874.
- Komorowska D. (2012) Organizacja produkcji i wyniki ekonomiczne gospodarstw ekologicznych o różnej wielkości. *Zeszyty Naukowe SGGW, Ekonomika i Organizacja Gospodarki Żywnościowej*, 95, Warszawa, 41-52.
- Krugman P., Wells R. (2012) *Mikroekonomia*. Wydawnictwo Naukowe PWN, Warszawa.
- Łuczka W. (2016) Stan badań nad rolnictwem ekologicznym w Polsce. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 453, 64-76.
- Patra S., Nayak S. R. (2012) A Theoretical Study on the Relationship between Wages and Labor Productivity in Industries. *International Journal of Economics and Research*, 157-163.

- Program Rozwoju Obszarów Wiejskich na lata 2014-2020 (PROW 2014-2020). Ministerstwo Rolnictwa i Rozwoju Wsi, Warszawa 2014.
- Rembisz W. (2007) Mikroekonomiczne podstawy wzrostu dochodów producentów rolnych. Vizja Press&IT, Warszawa.
- Rembisz W., Sielska A. (2014) Renta polityczna a inwestycje oraz relacje wynagrodzenia i wydajności czynnika pracy u producentów rolnych. [w:] Kowalski A., Wigier M., Wieliczko B. (red.) WPR a konkurencyjność polskiego i europejskiego sektora żywnościowego. Monografie Programu Wieloletniego 2011-2014, 146, IERiGŻ-PIB, Warszawa, 15-27.
- Rosenbaum P. R., Rubin D. B. (1983) The Central Role of the Propensity Score in Observational Studies for Causal effects. *Biometrika*, 70(1), 41-55.
- Sekhon J. S. (2011) Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R. *Journal of Statistical Software*, 42(7), 1-52.
- Sielska A., Kuszewski T., Pawłowska A., Bocian M. (2015) Wpływ polityki na kształtowanie się wartości dodanej. Monografie Programu Wieloletniego 2015-2019, 9, IERiGŻ-PIB, Warszawa.
- Sielska A., Pawłowska A. (2016) Szacowanie efektu oddziaływania polityki rolnej na wartość dodaną z wykorzystaniem propensity score matching. Monografie Programu Wieloletniego 2015-2019, 25, IERiGŻ-PIB, Warszawa.
- Strawiński P. (2014) Propensity Score Matching. Własności małopróbkowe. Wydawnictwo Uniwersytetu Warszawskiego.
- Szulc A. (2012) Ocena efektu oddziaływania: estymacja przez dopasowanie. [w:] Gruszczynski M. (red.) Mikroekonometria. Modele i metody analizy danych indywidualnych. Wydawnictwo Wolters Kluwer, Warszawa, 309-336.
- Trzciniński R. (2009) Wykorzystanie techniki propensity score matching w badaniach ewaluacyjnych. Polska Agencja Rozwoju Przedsiębiorczości, Warszawa.

#### **IMPACT OF INTERACTION BETWEEN SUBSIDIES ON AN INCREASE IN LABOUR PRODUCTIVITY IN POLISH FARMS**

**Abstract:** Labour productivity in agriculture occurs as one of the basic indicator in assessing the effectiveness of agriculture policy. The aim of research is to identify and quantify the treatment effect of subsidies on investment and environmental and LFA subsidies on the increase in labour productivity on Polish farms. The study uses data from the Farm Accountancy Data Network (FADN) database for individual Polish farms for 2007-2015. The applied research tool is Propensity Score Matching method.

**Keywords:** farms, labour productivity, agricultural policy, propensity score matching

## **EKSPERYMENTALNA OCENA EFEKTYWNOŚCI PORTFELA FUNDAMENTALNEGO DLA SPÓŁEK Z INDEKSU WIG20 ZA LATA 2004 – 2016**

**Michał Staszak**  
Wydział Zarządzania  
Uniwersytet Łódzki  
e-mail: [michal.staszak@gmail.com](mailto:michal.staszak@gmail.com)

**Streszczenie:** Przedmiotem badania było porównanie efektywności różnych metod konstrukcji portfeli fundamentalnych na przykładzie polskiego rynku kapitałowego. W tym celu wykorzystano klasyczną teorię portfelową oraz alternatywne podejście bazujące na taksonomicznej mierze atrakcyjności inwestycji (TMAI). Skuteczność obu metod poddano weryfikacji z wykorzystaniem spółek wchodzących w skład indeksu WIG20 w latach 2004 – 2016.

**Słowa kluczowe:** portfel fundamentalny, analiza portfelowa, TMAI, Markowitz, GPW

### **WPROWADZENIE**

Określenie składu portfela inwestycyjnego, optymalnego pod względem indywidualnych preferencji, jest podstawowym i najważniejszym elementem każdej skutecznej strategii inwestycyjnej. Jest to proces wieloetapowy, wymagający sprecyzowania wielu założeń i wyboru odpowiednich metod proces ten wspomagających. Zarówno w teorii, jak i praktyce rynków finansowych, znaleźć można cały szereg technik pozwalających na wyznaczenie rozwiązania tego problemu. Techniki te nie są jednak zbieżne, tzn. ich zastosowanie prowadzi zazwyczaj do różnych wyników.

Problematyka budowy portfeli inwestycyjnych była szeroko dyskutowana w licznych publikacjach naukowych. Za twórcę analizy portfelowej uważa się H. Markowitza, który zaproponował, aby dla minimalizacji ryzyka portfela o zadanej stopie zwrotu, dobierać do portfela instrumenty finansowe o najwyższej oczekiwanej stopie zwrotu, najniższej korelacji stóp zwrotu oraz najniższym

ryzyku. Ponadto, największe korzyści z punktu widzenia całego portfela można osiągnąć poprzez jego skuteczną dywersyfikację [Markowitz 1952]. Uproszczenie modelu Markowitza zaproponował w 1963 roku Sharpe, który bazował na założeniu, że zachowanie się cen poszczególnych akcji można opisać za pomocą jednego wskaźnika (współczynnik beta) charakteryzującego zachowanie się całego rynku [Sharpe1963]. Rozwinięciem jednowskaźnikowego modelu Sharpe'a jest model wyceny aktywów kapitałowych (Capital Assets Pricing Model - CAPM). Został on opracowany niezależnie przez Sharpe'a, Treynora, Lintnera oraz Mossina i implementuje powiązanie pomiędzy oczekiwaną stopą zwrotu z papierów wartościowych a ryzykiem, przy uwzględnieniu stopy wolnej od ryzyka [Fama, French 2004]. Alternatywne podejście do analizy portfelowej pokazali Łuniewska i Tarczyński [2006], Wskazali oni, że portfel inwestycyjny powinien składać się ze spółek charakteryzujących się dobrymi podstawami fundamentalnymi, które można opisać za pomocą analizy wskaźnikowej. W celu obiektywnej oceny kondycji fundamentalnej spółek Tarczyński [2002] zaproponował syntetyczną miarę TMAI (Taksonomiczna Miara Atrakcyjności Inwestycji), która umożliwiła sprowadzenie wielu kryteriów analizy fundamentalnej do jednej miary możliwej do wykorzystania w procesie budowy portfela.

W badaniu zaimplementowano metodę konstrukcji portfeli inwestycyjnych bazującą na wskaźniku TMAI [Wasilewska, Jasiukiewicz 2000], stanowiącym namiastkę zagregowanej siły fundamentalnej spółek [Tarczyńska-Łuniewska 2013], mierzoną za pomocą wskaźników ROE, ROA, P/E oraz P/BV. Analizę przeprowadzono na przykładzie spółek notowanych na GPW w Warszawie, wchodzących w skład indeksu WIG20 za okres 2004-2016. Efektywność badanych portfeli zweryfikowano poprzez porównanie ich wyników z wynikami portfeli referencyjnych. Jako benchmark przyjęto pasywne portfele replikujące zachowanie indeksu WIG20 oraz portfele skonstruowane w oparciu o klasyczną teorię Markowitza.

Celem pracy jest próba eksperymentalnego sprawdzenia, czy zastosowanie metod konstrukcji portfeli inwestycyjnych opartych na analizie siły fundamentalnej spółek prowadzi do ponadprzeciętnych zysków w długich okresach inwestowania, a generowane wyniki są lepsze niż wyniki osiągnięte z inwestycji w benchmark.

## DANE I METODY BADANIA

Na potrzeby tego badania zastosowano metody analizy portfelowej bazującej na wskaźniku TMAI do konstrukcji fundamentalnych portfeli inwestycyjnych. Do budowy portfeli wykorzystano spółki notowane na GPW w Warszawie w latach 2004 – 2016, wchodzące w skład indeksu WIG20. Jest to indeks grupujący 20 największych i najbardziej płynnych, a zarazem o największym pokryciu analitycznym, spółek notowanych na GPW. Pozwalało to przyjąć założenie, że wszelkie wyniki finansowe analizowanych spółek są natychmiast uwzględniane w wycenie, czyniąc je efektywnymi informacyjnie. Skład indeksu na każdy rok

ustalany był na bazie ostatniej rewizji indeksu z roku poprzedniego. Wykorzystano dane fundamentalne i rynkowe dla każdej ze spółek wchodzących do indeksu za cały okres badania z zastrzeżeniem, że w przypadku braku niezbędnych danych, spółka była pomijana w danym roku badania.

Dane finansowe pochodziły z rocznych sprawozdań skonsolidowanych lub jednostkowych (jeśli spółka nie tworzyła grupy kapitałowej) i obejmowały: zysk netto, kapitały własne oraz aktywa ogółem [Pomykalska, Pomykalski 2007]. Wskaźniki wyceny rynkowej wyznaczono korzystając z kapitalizacji rynkowej na koniec każdego z okresów. Z wykorzystaniem zgromadzonych danych przeprowadzono analizę wskaźnikową stosując podstawowe miary sytuacji finansowej przedsiębiorstwa oraz jego wyceny rynkowej: ROE, ROA, P/E i P/BV. Wskaźniki te wykorzystano następnie do skonstruowania miernika TMAI [Tarczyński 2002], na podstawie, którego nadano spółkom rangi odpowiadające ich sile fundamentalnej.

Fundamentalne portfele inwestycyjne budowano na koniec każdego analizowanego roku na podstawie danych z roku poprzedniego. W tym celu, w każdym z analizowanych okresów, wybierano pięć spółek o najwyższej wartości TMAI i nadawano im równe wagi udziału w portfelu. W symulacyjnym badaniu efektywności portfeli na rzeczywistych danych rynkowych przyjęto zasadę, że na koniec każdego okresu, akcje wchodzące w skład danego portfela są sprzedawane i całość uzyskanych środków jest inwestowana w nowy portfel spółek, wyznaczony na kolejny okres badania. W analizie tej nie uwzględniono podatków, kosztów transakcyjnych oraz niepodzielności akcji.

Dla oceny przydatności portfeli fundamentalnych wykorzystano portfele benchmarkowe bazujące na teorii Markowitza. Podejście to różni się zasadniczo od metody opartej na wskaźniku TMAI, gdyż wykorzystuje wyłącznie parametry charakteryzujących stopy zwrotu z inwestycji, a nie *explicite* fundamenty spółek. Jednakże jako jedna z podstawowych technik konstrukcji portfeli, została wybrana jako punkt odniesienia do porównania efektywności analizowanych portfeli. Zbadano trzy różne rodzaje portfeli inwestycyjnych zbudowanych wg klasycznej teorii portfelowej Markowitza [1952]:

- portfel o minimalnym ryzyku (bez ograniczeń dotyczących krótkiej sprzedaży),
- portfel o minimalnym ryzyku (z restrykcjami dotyczącymi krótkiej sprzedaży),
- portfel rynkowy (o maksymalnym stosunku zysku do ryzyka).

Każdy z portfeli budowano na koniec danego roku na podstawie dziennych stóp zwrotu za rok poprzedni. Przyjęto, że oczekiwana dzienna stopa zwrotu równa jest średniej dziennych stóp zwrotu z poprzedniego okresu, a oczekiwana korelacja równa jest historycznej korelacji dziennych stóp zwrotu z roku poprzedniego. Do wyznaczenia wag portfela rynkowego założono, że stopa wolna od ryzyka równa jest średniej rocznej rentowności obligacji 10-cio letnich Skarbu Państwa.

Badanie zakończono porównaniem efektywności analizowanych portfeli fundamentalnych z benchmarkami. W tym celu stopy zwrotu generowane z portfeli



TMAI porównywano ze stopą zwrotu z indeksu WIG20 oraz z portfeli Markowitza w ten sposób, że dla portfela wyznaczonego na bazie TMAI na koniec roku 2014 generowano stopę zwrotu za rok 2015 i porównano z wynikiem osiągniętym w tym roku przez benchmarki, dla portfela fundamentalnego skomponowanego na danych za rok 2015 porównano stopę zwrotu wygenerowaną za rok 2016 ze stopami z benchmarków itd. Otrzymane wyniki roczne i skumulowane poddano dyskusji w aspekcie przydatności i korzyści płynących z zastosowanych metod konstrukcji i zarządzania portfelami inwestycyjnymi.

## WYNIKI

Portfel fundamentalny, zbudowany ze spółek z najwyższą wartością wskaźnika TMAI, osiągnął w ciągu analizowanych 13 lat łączną stopę zwrotu w wysokości 143,6%, podczas gdy WIG20 wypracował w tym samym okresie stopę zwrotu na poziomie 23,3%. W latach 2004-2016, portfel fundamentalny okazał się gorszy od indeksu WIG20 tylko czterokrotnie: w 2005, 2007, 2008 oraz w 2015 roku, przy czym największa ujemna różnica wyniosła 13,4 pkt proc. W pozostałych okresach portfel osiągał wyższe stopy zwrotu od indeksu, a największą różnicę odnotował w roku 2009 – 69,6%. Łączna różnica stóp zwrotu w okresie badania wyniosła 120,2 pkt proc. na korzyść portfela fundamentalnego.

Portfele zbudowane na bazie teorii Markowitza wygenerowały skumulowaną stopę zwrotu wyższą niż indeks WIG20 na przestrzeni analizowanych 13 lat. Najlepiej zachował się portfel z restrykcjami dotyczącymi krótkiej sprzedaży. Wygenerował skumulowaną stopę zwrotu na poziomie 70,8%. Zbliżony wynik osiągnął portfel bez restrykcji – w tym przypadku stopa zwrotu wyniosła 70,0%. Najgorzej w tym zestawieniu zachował się portfel rynkowy, którego łączna skumulowana stopa zwrotu wyniosła 59,9%.

Analizując jednookresowe (tabela 1) stopy zwrotu, warto zwrócić uwagę na fakt, że oba portfele o minimalnym ryzyku zachowały się lepiej od benchmarku w roku 2008, kiedy na rynkach panował kryzys finansowy, a WIG20 zanotował spadek o 48,2%. Najgorszą roczną stopę zwrotu w porównaniu do WIG20 zanotował w roku 2009 portfel rynkowy – przy wzroście indeksu o ponad 33%, wartość portfela spadła o 9,7%, skutkując ujemną różnicą stóp zwrotu w wysokości 43,2 pkt proc. Jednocześnie portfel rynkowy osiągnął najlepszy jednoroczny wynik w stosunku do WIG20 w roku 2006 - wygenerowana stopa zwrotu była wyższa od indeksu o 40,2%. Taka rozpiętość wyników może być tłumaczona dużą koncentracją portfela – był to portfel zdecydowanie najmniej zdywersyfikowany w porównaniu z portfelami o minimalnym ryzyku. Biorąc pod uwagę liczbę lat, kiedy dane portfele były lepsze lub gorsze od indeksu WIG20, nie można wskazać jednoznacznie portfela charakteryzującego się największą stabilnością i konsekwencją wyników.

Tabela 1. Jednookresowe i skumulowane stopy zwrotu z portfeli w latach 2004-2016

Stopy zwrotu	Portfel TMAI	Portfel o minimalnym ryzyku (bez restrykcji)	Portfel o minimalnym ryzyku (z restrykcjami)	Portfel rynkowy	WIG20
2004	32,5%	9,8%	9,9%	-12,2%	24,1%
2005	34,1%	4,3%	11,4%	66,1%	35,4%
2006	37,8%	51,3%	47,3%	63,9%	23,7%
2007	3,3%	23,8%	20,8%	24,2%	5,2%
2008	-61,6%	-37,6%	-43,2%	-53,6%	-48,2%
2009	103,1%	30,2%	31,9%	-9,7%	33,5%
2010	20,0%	4,3%	6,4%	54,9%	14,9%
2011	-5,1%	10,6%	6,4%	-27,5%	-21,9%
2012	23,1%	6,7%	11,6%	7,6%	20,4%
2013	-6,2%	4,5%	1,8%	-7,6%	-7,0%
2014	7,0%	1,4%	-2,7%	0,8%	-3,5%
2015	-23,0%	-23,8%	-16,7%	3,0%	-19,7%
2016	14,1%	-1,8%	0,4%	10,9%	4,8%
Łącznie	143,6%	70,0%	70,8%	59,9%	23,3%

Źródło: opracowanie własne

Porównując skumulowane stopy zwrotu wszystkich analizowanych portfeli, należy zauważyć, że każdy z nich na przestrzeni 13 lat osiągnął wyniki lepsze niż indeks WIG20, a najbardziej dochodowa okazała się strategia inwestycyjna oparta na wskaźniku TMAI. Inwestowanie w 5 najlepszych spółek według rankingu siły fundamentalnej pozwoliło uzyskać wyniki wyższe od benchmarków. Nie pozwala to jednak na stwierdzenie, że inwestycja w portfele Markowitza jest gorszą strategią inwestycyjną niż inwestycja w portfel TMAI, gdyż w powyższych analizach jedyną porównywaną miarą była stopa zwrotu opisująca dochodowość portfeli, a pominięty został aspekt ryzyka, którego minimalizacja stoi u podstaw klasycznej teorii portfelowej.

Powyższe wyniki potwierdzają słuszność przyjętych założeń i skuteczność zastosowanej metody, mimo ograniczonego (do zaledwie 4-ch) katalogu wykorzystanych wskaźników fundamentalnych. Uprawniony wydaje się zatem wniosek, że inwestowanie w spółki o najlepszej kondycji finansowej i najkorzystniejszej wycenie rynkowej jest rozwiązaniem z punktu widzenia inwestora lepszym, niż pasywne inwestowanie we wszystkie spółki z indeksu. Wymaga to poświęcenia dużej ilości czasu na analizę spółek oraz budowę i zarządzanie portfelem, ale osiągnięte wyniki jednoznacznie pokazują wymierne korzyści takiego podejścia do inwestowania.

## WNIOSKI

Prezentowane w niniejszym badaniu metody konstrukcji portfeli inwestycyjnych okazały się przydatnymi narzędziami w procesie inwestowania kapitału na GPW. Zarówno portfele oparte na wskaźniku siły fundamentalnej, jak i portfele zbudowane na bazie teorii Markowitza wygenerowały stopy zwrotu wyższe niż indeks WIG20, co potwierdza ich skuteczność w praktycznym zastosowaniu.

W konsekwencji uprawnione wydaje się twierdzenia, że zastosowanie na polskim rynku kapitałowym metod konstrukcji portfeli fundamentalnych, prowadzi do ponadprzeciętnych zysków w długich okresach inwestowania. Wyniki te są lepsze niż wyniki osiągnięte z inwestycji w benchmark. Przeprowadzone badanie obejmowało 13 lat o krańcowo zmiennych i turbulentnych warunkach koniunktury na rynkach finansowych, ze światowym kryzysem finansowym włącznie. Był to również okres silnej globalizacji i rozwoju rynków kapitałowych. Wyniki badania dowodzą, że prezentowane metody budowy portfeli są aktualne bez względu na panujące trendy czy otoczenie rynkowe i dokonujące się w nim zmiany. W długim horyzoncie czasowym inwestowanie oparte o analizę portfelową wydaje się więc być kluczowym czynnikiem determinującym sukces każdego inwestora giełdowego.

## BIBLIOGRAFIA

- Chorkowy B., Drymluch M. (2008) Wielowymiarowa analiza porównawcza banków notowanych na Giełdzie Papierów Wartościowych w Warszawie. [w:] Kopycińska D. Konkurencyjność podmiotów rynkowych. Wydawnictwo Katedry Mikroekonomii Uniwersytetu Szczecińskiego, 54-64.
- Fama E. F., French K. R. (2004) The Capital Asset Pricing Model: Theory and Evidence. *Journal of Economic Perspectives*, 18(3), 25-46.
- Łuniewska M., Tarczyński W. (2006) Metody wielowymiarowej analizy porównawczej na rynku kapitałowym. Wydawnictwo Naukowe PWN, Warszawa.
- Markowitz H. (1952) Portfolio Selection. *The Journal of Finance*, 1(7), 77-91.
- Sharpe W. F. (1963) A Simplified Model for Portfolio Analysis. *Management Science*, 9(2), 277-293.
- Tarczyński W. (2002) Fundamentalny portfel papierów wartościowych. PWE, Warszawa.
- Tarczyński W. (2017) Fundamentalny portfel papierów wartościowych jako alternatywa dla modelu Markowitza. [w:] Ubezpieczenia i finanse. Rozwój i Perspektywy. Wydawnictwo Uniwersytetu Łódzkiego, 187-198.
- Tarczyńska-Łuniewska M. (2013), Metodologia oceny siły fundamentalnej spółek (giełdowych i pozagiełdowych). ZAPOL, Szczecin.
- Pomykańska B., Pomykański P. (2007) Analiza finansowa przedsiębiorstwa. Wydawnictwo Naukowe PWN.

Wasilewska H., Jasiukiewicz M. (2000) Taksonomiczna miara atrakcyjności inwestycji w akcje na przykładzie wybranych spółek giełdowych. *Bank i Kredyt*, 6, 78-87.

**EXPERIMENTAL ASSESSMENT OF FUNDAMENTAL  
PORTFOLIO'S EFFECTIVENESS BASED ON STOCKS INCLUDED  
IN WIG20 INDEX IN THE PERIOD 2004-2016**

**Abstract:** The aim of this paper is a comparison of effectiveness of different methods related to construction of fundamental portfolios in the case of polish capital market. The study describes classical portfolio theory and alternative approach based on the Taksonomiczna Miara Atrakcyjności Inwestycji (TMAI) measure. The effectiveness of all methods is verified using the companies included in the WIG20 index in the period 2004 - 2016.

**Keywords:** fundamental portfolio, portfolio analysis, TMAI, Markowitz, WSE

## IDENTYFIKACJA KOMPONENTÓW DESTRUKCYJNYCH W MODELACH PREDYKCYJNYCH W PODEJŚCIU WIELOMODELOWYM

**Ryszard Szupiluk, Paweł Rubach**

Kolegium Analiz Ekonomicznych

Szkoła Główna Handlowa w Warszawie

e-mail: ryszard.szupiluk@sgh.waw.pl; pawel.rubach@sgh.waw.pl

**Streszczenie:** W niniejszym artykule przedstawimy metodę identyfikacji komponentów destrukcyjnych występujących w podejściu wielomodelowym wykorzystującym algorytmy ślepej separacji sygnałów. Ocena charakterystyki poszczególnych komponentów dokonana zostanie na podstawie autorskich mierników zmienności/gładkości sygnałów. W celu potwierdzenia skuteczności prezentowanej metody przedstawimy praktyczny eksperyment poprawy wyników prognozy zużycia energii elektrycznej.

**Słowa kluczowe:** predykcja, agregacja modeli, ślepa separacja, identyfikacja szumów

### WPROWADZENIE

Rosnąca popularność podejścia wielomodelowego jest naturalną konsekwencją rozwoju metod tworzenia modeli w takich dyscyplinach jak Data Mining, uczenie maszynowe, sztuczna inteligencja czy modelownie nieparametryczne. Dojrzałość podstawowych metod predykcji powoduje, że dalszej poprawy jakości predykcji upatrywać można w działaniach prowadzących do połączenia wiedzy pochodzącej z różnych modeli ujmujących badane zjawisko w odmienny sposób. Jest wiele popularnych metod agregacji, z których większość oparta jest na różnych formach uśredniania wyników lub parametrów modeli [Clements 1989, Gatnar 2008]. Jednym z zasadniczych ograniczeń takich metod jak boosting lub bagging jest zawężenie agregacji do modelu w zasadzie tego samego typu, gdzie zróżnicowanie wiąże się przede wszystkim z uczeniem na różnych podzbiorach [Breiman 1996]. W istotny sposób ogranicza to ideę agregacji różnych modeli jako koncepcję łączenia różnych źródeł informacji.

Tych ograniczeń nie posiada podejście agregujące oparte na metodach ślepej separacji [Szupiluk i in. 2004, Szupiluk 2013]. W podejściu tym, traktujemy zbiór rezultatów predykcji generowanych przez różne modele, jako jedną wielowymiarową zmienną. Zakładamy, że zmienna ta zawiera ukryte komponenty konstruktywne i destrukcyjne dla wyniku prognozy. Oczekujemy przy tym, że owe ukryte komponenty bazowe mają pewien „fizyczny charakter”, zaś eliminacja komponentów destrukcyjnych powinna skutkować poprawą prognozy mierzoną różnymi kryteriami. Ten aspekt eliminacji „niejako” fizycznych zakłóceń zidentyfikowanych metodami ślepej separacji sygnałów (ang. Blind Signal Separation BSS) odróżnia prezentowane podejście od agregacji opartej na optymalizacji wybranego pojedynczego kryterium błędu.

W niniejszym artykule skoncentrujemy się na aspekcie identyfikacji komponentów destrukcyjnych. Jest to zagadnienie o tyle złożone, że komponenty owe mają zwykle charakterystykę statystyczną znacząco odmienną od klasycznych białych szumów – często przyjmowanych jako elementy zakłócające. Powoduje to konieczność poszukiwania nowych miar i charakterystyk – w stosunku do standardowych metod opartych na metodach korelacyjnych, R/S lub spektralnych [Hurst 1951, Vasegi 1997, Therrien 1992]. W niniejszym artykule rozwiniemy klasę miar zmienności mogących stanowić alternatywę lub uzupełnienie dla owych klasycznych metod. Całokształt rozważań zostanie zobrazowany praktycznym przykładem prognozowania energii.

## AGREGACJA BSS

Załóżmy istnienie zbioru  $m$  rezultatów predykcji  $x_i$ ,  $i = 1, \dots, m$ . Wyniki te zbierzemy w jednej wielowymiarowej zmiennej  $\mathbf{x} = [x_1, \dots, x_m]^T$ . Przyjmijmy dalej, że wyniki są akceptowalnej jakości co oznacza, iż prognozowane wartości w pewnym stopniu odpowiadają wartościom obserwowanym a w pewnym się różnią. Możemy więc powiedzieć, że dany rezultat jest kombinacją komponentów konstruktywnych  $\hat{s}_j$ ,  $j = 1, \dots, p$  związanych z podobieństwem oraz destrukcyjnych  $\tilde{s}_l$ ,  $l = 1, \dots, q$  związanych z różnicami wartości przewidywanych i obserwowanych. Wszystkie te komponenty potraktujemy jako ukryte komponenty bazowe zawarte w wielowymiarowej zmiennej  $\mathbf{x}$ . W przypadku liniowego sposobu (systemu) mieszania można więc zapisać

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k), \quad (1)$$

gdzie  $k$  oznacza numer obserwacji lub indeks czasu, macierz  $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{m \times n}$  reprezentuje system mieszający zaś wektor  $\mathbf{s}(k) = [\hat{s}_1(k), \dots, \hat{s}_p(k), \tilde{s}_{p+1}(k), \dots, \tilde{s}_{p+q}(k)]^T$  reprezentuje zbiór komponentów bazowych, gdzie  $n = p + q$ . Dla uproszczenia przyjmijmy, że  $m = n$ , macierz  $\mathbf{A}$  jest nieosobliwa, zaś  $E\{\mathbf{s}\} = \mathbf{0}$ . Zauważmy dalej,

że identyfikując system mieszający  $\mathbf{A}$  oraz komponenty bazowe  $\mathbf{s}$  oraz eliminując komponenty destrukcyjne (stawiając odpowiednio  $\tilde{s}_i = 0$ ) otrzymamy

$$\hat{\mathbf{x}}(k) = \mathbf{A}[\hat{s}_1(k), \dots, \hat{s}_p(k), 0_{p+1}(k), \dots, 0_n(k)]^T, \quad (2)$$

gdzie  $\hat{\mathbf{x}} = [\hat{x}_1, \dots, \hat{x}_m]^T$  jest „poprawianą” wersją rezultatów predykcji  $\mathbf{x}$ . Kluczowym zagadnieniem w rozważanej procedurze, jest znalezienie takiej transformacji, która rzeczywiście prowadzi do dekompozycji danych  $\mathbf{x}$  na komponenty konstruktywne i destrukcyjne. A następnie poprawne określenie które komponenty są destrukcyjne.

Punktem wyjścia w poszukiwaniu komponentów bazowych może być oczekiwanie, że mają one pewien „fizyczny” charakter związany z czynnikami determinującymi proces modelowania. Na cały proces możemy także spojrzeć jak na próbę separacji nieznanymi komponentów bazowych zmieszanych w nieznanym systemie. Prowadzi to wprost do tzw. problemu ślepej separacji i rozwiązań w nim stosowanych [Comon i in. 2010, Hyvarinien i in. 2001], co można przedstawić jako poszukiwanie transformacji odwrotnej do (1) takiej, że

$$\mathbf{y}(k) = \mathbf{W}\mathbf{x}(k) \approx \mathbf{s}(k), \quad (3)$$

gdzie macierz  $\mathbf{W} = \mathbf{A}^{-1}$  pełni rolę systemu separującego odwrotnego do  $\mathbf{A}$ .

Istnieje wiele metod BSS znajdowania  $\mathbf{W}$  takich jak algorytmy analizy składowych niezależnych (ang. Independent Component Analysis-ICA) czy też algorytmy AMUSE lub SOBI [Common i in. 2010, Hyvarinien i in 2001, Szupiluk 2013]. Znaczną część tych rozwiązań można sprowadzić, rozszerzyć lub uogólnić do przypadku eksploracji uogólnionej macierzy kowariancji postaci

$$\mathbf{R}_F = E \left\{ \left( \lambda \mathbf{y}(k) + \beta \mathbf{f}(\mathbf{y}(k)) \right) \left( \sum_{j \in J} a_j \mathbf{y}(k-j) \right)^T \right\}, \quad (4)$$

gdzie  $\lambda, \beta$  są parametrami określającymi proporcje między macierzą kowariancji oraz nieliniową macierzą kowariancji,  $J$  jest zbiorem wartości opóźnień pozwalających uwzględnić strukturę czasową sygnałów. Eksploracja macierzy (4) łączy w sobie charakterystyki wykorzystywane zarówno przez algorytmy ICA bazujące na statystykach wyższych rzędów jak również charakterystyki wykorzystywane przez algorytmy typu BSS – SOS (ang. Second Order Statistics) eksplorujące strukturę czasową sygnałów. Zauważmy, że dla  $\lambda = 1$   $\beta = 0$  oraz

$J = \{0\}$  otrzymujemy zwykłą macierz kowariancji  $\mathbf{R}_F = E \left\{ \mathbf{y}\mathbf{y}^T \right\}$ , dla  $\lambda = 1$ ,  $\beta = 0$ ,  $J \neq \{0\}$  otrzymujemy macierz kowariancji z opóźnieniami

$\mathbf{R}_F = E \left\{ \mathbf{y}(k) \left( \sum_{j \in J} a_j \mathbf{y}(k-j) \right)^T \right\}$  zaś dla  $\lambda = 0$   $\beta = 1$ ,  $J = \{0\}$ , otrzymujemy

nieliniową macierz kowariancji  $\mathbf{R}_F = E\{\mathbf{f}(\mathbf{y})\mathbf{y}^T\}$ . Zasadniczą kwestią jest dobór nieliniowości występujących w (4). Teoretycznie jej optymalna postać dla algorytmów ICA ma postać  $f(y) = -\partial \log(p(y)) / \partial y$  gdzie  $p(y)$  jest funkcją gęstości prawdopodobieństwa zmiennej  $y$ .

Ze względu na ograniczoną informację a priori co do postaci separowanych sygnałów jako podstawa wyprowadzenia nieliniowości proponowane są najczęściej ogólne modele rozkładów pozwalające na elastyczne dopasowanie funkcji rozkładu do przetwarzanych danych.

Uwzględniając powyższe zależności można podać algorytm separacji postaci

1. Niech  $\mathbf{z}(t) = \mathbf{x}(t)$ ,  $\mathbf{W} = \mathbf{I}$
2. Oblicz generalną postać macierzy korelacji opóźnień  $\mathbf{R}_F$  zgodnie z wzorem (4)
3. Znajdź macierz ortogonalną  $\mathbf{Q}_p$  (np. za pomocą SVD), która diagonalizuje  $\mathbf{R}_F + \mathbf{R}_F^T$  i wykonaj dekorelację  $\mathbf{y}(t) = \mathbf{Q}_p \mathbf{z}(t)$
4. Niech  $\mathbf{W} \leftarrow \mathbf{Q}_p \mathbf{W}$ ,  $\mathbf{z}(t) \leftarrow \mathbf{y}(t)$  i powróć do kroku nr 2 dopóki nie zostanie osiągnięte zamierzone kryterium sukcesu.

Jak wcześniej powiedziano w zależności od parametrów macierzy  $\mathbf{R}_F$  otrzymujemy rodzaj eksplorowanej charakterystyki. Ma to istotny wpływ na operacjonalizację powyższego algorytmu. W przypadku dominacji statystyk drugiego rzędu liczbę iteracji można ograniczyć nawet do dwóch, co w wielu przypadkach wystarcza do efektywnej separacji. Ograniczeniem jest jednak charakterystyka separowanych sygnałów, muszą one mieć strukturę czasową o różnych funkcjach autokorelacji. Ograniczenia tego nie posiada eksploracja zależności nieliniowych, jednak w tym wypadku liczba iteracji musi być znacznie większa. Efektywne działanie algorytmu łączącego jednocześnie aspekty liniowości, nieliniowości oraz opóźnień jest otwartym i złożonym zagadnieniem.

## IDENTYFIKACJA KOMPONENTÓW DESTRUKCYJNYCH

Przedstawiony powyżej schemat konceptualny definiujący proces łączenia modeli przy jednoczesnej redukcji zakłóceń, jest schematem ideowym. Zakładając jego pełną realizację moglibyśmy oczekiwać, że wydzielone komponenty destrukcyjne będą związane z konkretnymi czynnikami zniekształcającymi wyniki predykcji takie jak zakłócenia i szумы w danych uczących, nie w pełni adekwatny dobór modeli predykcyjnych czy niedokładności procesu optymalizacji. W efekcie eliminacja takiego „fizycznego” komponentu powinna skutkować poprawą predykcji na wszelkich kryteriach oceny. Owa motywacja przyświeca oparciu etapu dekompozycji na metodach ślepej separacji. Oczywiście w rzeczywistości ów czysty schemat ideowy rzadko może być w pełni zrealizowany.



W praktyce założenia co do statycznej i liniowej postaci modelu mieszającego oraz wybór określonej metody separacji np. ICA mogą być nie w pełni adekwatne. Rzeczywiste komponenty mogą być zmieszane w sposób nieliniowy lub dynamiczny zaś dla jego identyfikacji mogą być właściwe metody bazujące na nieujemności, gładkości lub rzadkości (ang. sparse) sygnałów. Należy jednak zaznaczyć, że te różnice choć mogące być istotnymi dla konkretnego przypadku nie zmieniają ogólnych ram naszego schematu badawczego. Ponieważ problematyka rozwiązywania nieliniowych oraz dynamicznych problemów BSS nadal jest w dużym stopniu otwarta, to istniejące efektywne algorytmy separacji dla przypadku liniowego, są stosowane w każdym typie modelu mieszającego. W kontekście metod ślepej separacji istotnym elementem jest także liczba sygnałów użytych w procesie mieszania/separacji, tu także nie dysponujemy wiedzą ile i jakiego typu modeli predykcyjnych należałoby użyć dla uzyskania idealnych komponentów konstrukcyjnych i destrukcyjnych. Wymiar przyjmowany jest zgodny z liczbą posiadanych modeli.

Wobec powyższych uwarunkowań można przyjąć, że choć celem dekompozycji wyników predykcji za pomocą metod ślepej separacji jest identyfikacja czystych komponentów destrukcyjnych, to w praktyce otrzymujemy najczęściej tylko pewien analityczny zbiór komponentów bazowych, których zarówno charakterystyki matematyczne jak i ich wpływ na wyniki predykcji nie są w pełni jednoznaczne. Owa niejednoznaczność oraz trudność oceny w dużej mierze wiąże się z właściwościami i charakterystyką standardowych metod i miar używanych do oceny zaszumienia i losowości sygnałów. Analiza wariancji, korelacji, autokorelacji, spektralna lub analiza R/S są efektywnymi narzędziami w przypadkach standardowych szumów modelowanych jako matematyczne białe szumy [Hamilton 1994, Shirayev 1999, Therrien 1992]. W szczególności odnosi się to do sygnałów gausowskich które są w pełni statystycznie określone statystykami drugiego rzędu. Niestety w przypadkach z jakimi mamy do czynienia przy dekompozycjach rzeczywistych sygnałów takich jak np. zużycie energii elektrycznej owe standardowe metody nie są w pełni skuteczne. Stąd potrzeba opracowań nowych koncepcji oceny zmienności i losowości sygnałów. Istnieją różne koncepcje takiej oceny [Szupiluk i inni 2010], obecnie rozwiniemy metodę opartą na zestawie miar zmienności.

Punktem wyjścia naszej koncepcji będzie powiązanie zakłóceń z szeroko rozumianą zmiennością (możliwe są tu także nawiązania do gładkości, rozproszenia, rozrzutu) dla której można zaproponować następującą miarę:

$$\Theta_1(y) = \frac{1}{N} \frac{\sum_{k=2}^N |y(k) - y(k-1)|}{I(\max(y) - \min(y))}, \quad (5)$$

gdzie:  $I(u) = u$  dla  $u \neq 0$  oraz  $I(u) = 1$  dla  $u = 0$ .

Interpretacja miary (5) jest stosunkowo prosta i intuicyjna. Otrzymujemy średnią bezwzględną (całkowitą) zmienność sygnału w stosunku do zakresu wahań. Funkcja  $I(u)$  pełni rolę indykatora zera i zapobiega dzieleniu przez zero dla sygnału stałego. Formuła (5) wartość maksymalną równą 1 gdy wszystkie zmiany sygnału równe są zakresowi wahań, przyjmuje zaś wartość minimalną równą 0 dla sygnału stałego. Miara (5) i jej interpretacja stanowią punkt wyjścia dla jej zmodyfikowanych wersji. Przede wszystkim w zastosowaniach optymalizacyjnych w miejsce czasem kłopotliwej wartości bezwzględnej można zastosować formułę postaci

$$\Theta_2(y) = \frac{1}{N} \frac{\sum_{k=2}^N (y(k) - y(k-1))^2}{I((\max(y) - \min(y))^2)}, \quad (6)$$

lub

$$\Theta_3(y) = \frac{1}{N} \frac{\sum_{k=2}^N \log(\cosh(y(k) - y(k-1)))}{I(\log(\cosh(\max(y) - \min(y))))}. \quad (7)$$

Powyższe miary (5)-(7) oprócz bezpośredniej miary zmienności/gładkości w różnych postaciach analitycznych można wiązać także z innymi charakterystykami. Wyrażenie (5) można także interpretować w kategoriach rzadkości (ang. sparse) sygnałów. Z kolei formułę (6) można powiązać z analizą R/S na bazie eksploracji ilorazu statystyki drugiego rzędu i rozstępu. Formuła (7) ze względu na popularną w obszarze machine learning funkcję  $\log(\cosh(y))$  może być interesująco interpretowana w kategoriach statystyk wyższych rzędów [Oja 1995].

W efekcie, choć każda z powyższych miar może być wykorzystana do oceny zmienności/gładkości, to głębsze praktyczne badanie charakterystyki sygnałów dokonane powinno być raczej z łącznym użyciem powyższych miar. W tym celu możemy rozważyć charakterystyki sygnałów w pewnej przestrzeni określonej wymiarami  $\Theta_1(y)$ ,  $\Theta_2(y)$ ,  $\Theta_3(y)$ . Ze względu na łatwość graficznej inspekcji wygodnym może być rozważanie par charakterystyk w układzie kartezjańskim, co np. dla  $\Theta_1(y)$ ,  $\Theta_2(y)$  prowadzi do możliwości intuicyjnej miary łącznej zmienności postaci

$$\Theta_E(y) = \sqrt{(\Theta_1(y))^2 + (\Theta_2(y))^2}. \quad (8)$$

Miara (8) określa odległość euklidesową sygnału od środka układu określonego przez charakterystyki  $\Theta_1(y)$ ,  $\Theta_2(y)$  i może być rozszerzona na większą liczbę zmiennych.

## EKSPERYMENT PRAKTYCZNY

W niniejszym eksperymencie przedstawimy skuteczność prezentowanej metody dla poprawy predykcji krótkoterminowego zużycia energii elektrycznej. W badaniu wykorzystano godzinowe dane o łącznym zużyciu energii pochodzące z polskiego systemu energetycznego z lat 1987-1996. Dla agregacji wykorzystano sześć modeli neuronowych typu MLP optymalizowanych algorytmem Levenberga-Marquarda. Do agregacji wykorzystano próbę 2000 obserwacji. W efekcie uzyskano prognozy o następujących wielkościach błędu MSE oraz MAPE (tabela 1).

Tabela 1. Błędy prognozy dla poszczególnych modeli neuronowych

	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>
MSE (10 <sup>-4</sup> )	8,5616	8,3066	6,9451	8,2068	6,2407	8,8686
MAPE (10 <sup>-2</sup> )	2,1454	2,1408	1,9121	2,1108	1,8243	2,1855

Źródło: opracowanie własne

W kolejnym kroku dokonano dekompozycji prognoz na komponenty bazowe  $y_1 \dots y_6$  za pomocą algorytmu AMUSE zgodnie z równaniem (3). Następnie podjęto próbę identyfikacji komponentów konstruktywnych i destruktywnych przy użyciu badania odległości euklidesowej pomiędzy wektorami wartości prognoz poszczególnych komponentów. Wyniki zebrane w tabeli 2 nie pozwalają jednak na jednoznaczną kategoryzację komponentów. Można jedynie zauważyć, że od pozostałych komponentów wyraźnie odróżnia się komponent  $y_1$ .

Tabela 2. Odległości euklidesowe między wektorami komponentów bazowych

	y <sub>1</sub>	y <sub>2</sub>	y <sub>3</sub>	y <sub>4</sub>	y <sub>5</sub>	y <sub>6</sub>
y <sub>1</sub>	0,00	25 693,78	25 700,56	25 579,62	25 656,51	25 679,86
y <sub>2</sub>	25 693,78	0,00	3 639,29	2 989,31	3 335,86	3 730,19
y <sub>3</sub>	25 700,56	3 639,29	0,00	3 609,59	3 577,76	3 919,25
y <sub>4</sub>	25 579,62	2 989,31	3 609,59	0,00	3 276,68	3 415,62
y <sub>5</sub>	25 656,51	3 335,86	3 577,76	3 276,68	0,00	3 675,32
y <sub>6</sub>	25 679,86	3 730,19	3 919,25	3 415,62	3 675,32	0,00

Źródło: opracowanie własne

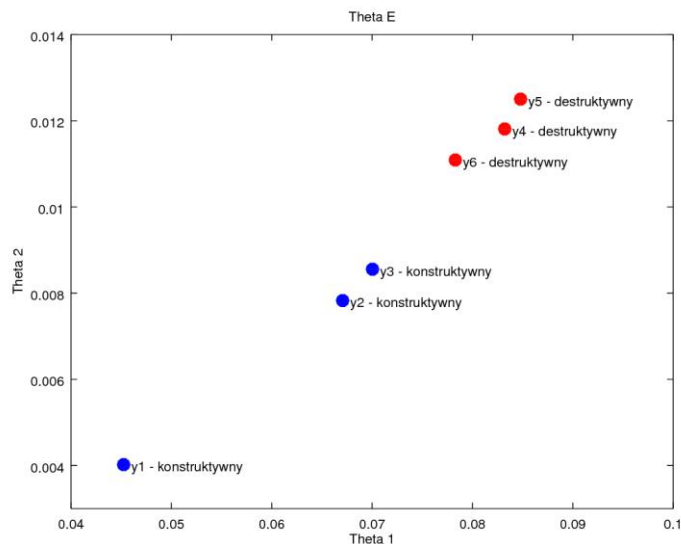
W następnym etapie eksperymentu dokonano obliczeń wariancji, współczynników Hursta oraz miar zmienności  $\Theta_1(y)$ ,  $\Theta_2(y)$ ,  $\Theta_3(y)$  oraz  $\Theta_E(y)$  zaproponowanych w niniejszej pracy. Wyniki przedstawiono w tabeli 3.

Tabela 3. Wariancja, współczynnik Hursta oraz miary zmienności  $\Theta(y)$ 

	$y_1$	$y_2$	$y_3$	$y_4$	$y_5$	$y_6$
$D^2(y)$	0,0134	1,3331	1,4180	1,4616	1,4896	1,7895
Hurst	0,8700	0,6400	0,6300	0,6300	0,7000	0,5300
$\Theta_1(y)$	0,0452	0,0670	0,0700	0,0832	0,0848	0,0783
$\Theta_2(y)$	0,0040	0,0078	0,0086	0,0118	0,0125	0,0111
$\Theta_3(y)$	0,0042	0,0293	0,0332	0,0410	0,0430	0,0442
$\Theta_E(y)$	0,0454	0,0675	0,0706	0,0840	0,0857	0,0791

Źródło: opracowanie własne

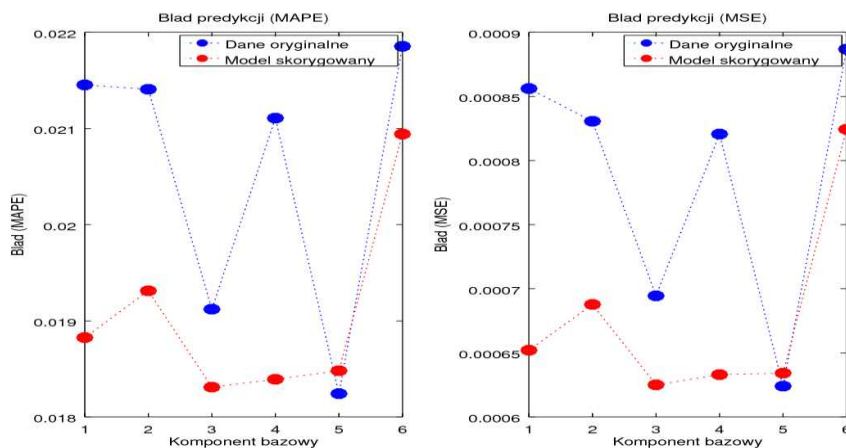
Jak wynika z tabeli 3 pierwszy z komponentów  $y_1$  z uwagi na niską wariancję oraz wysoki współczynnik Hursta należy najprawdopodobniej do komponentów konstruktywnych. Z drugiej jednak strony, wartości wariancji oraz współczynnika Hursta nie pozwalają na odróżnienie komponentów destrukcyjnych i konstrukcyjnych spośród pozostałych pięciu komponentów  $y_2 \dots y_6$ . Można natomiast analizując wartości miar zmienności dojść do wniosku, że destrukcyjne w tym przypadku są ostatnie trzy komponenty  $y_4, y_5, y_6$ , gdyż w ich przypadku wartości miary  $\Theta_E(y)$  są istotnie wyższe niż w przypadku pierwszych trzech komponentów (powyżej 0,79) co pokazuje rysunek 1, w którym miara jest widoczna jako odległość poszczególnych punktów od punktu (0,0) na wykresie.

Rysunek 1. Miary zmienności  $\Theta_1(y)$ ,  $\Theta_2(y)$  oraz  $\Theta_E(y)$ 

Źródło: opracowanie własne

Na podstawie powyższej obserwacji dokonano eliminacji 3 komponentów bazowych  $y_4$ ,  $y_5$ ,  $y_6$  w wyniku, którego średni błąd predykcji zużycia energii elektrycznej zmniejszył się średnio o 7,8% jeśli zmierzmy go za pomocą metody MSE oraz 3,8% w przypadku MAPE. Rysunek 2 ilustruje szczegółowe zmiany błędów predykcji dla poszczególnych komponentów bazowych.

Rysunek 2. Błędy predykcji przed i po eliminacji komponentów destrukcyjnych



Źródło: opracowanie własne

## PODSUMOWANIE

Przedstawiona koncepcja identyfikacji komponentów destrukcyjnych opiera się na nowych charakterystykach zmienności pozwalających dokonywać stosunkowo precyzyjnych pomiarów wpływu poszczególnych komponentów na wyniki prognoz. Zagadnienie identyfikacji zakłóceń/szumów przedstawione zostało w ramach predykcyjnego podejścia wielomodelowego, jednak prezentowana metoda może zostać zastosowana w dowolnym problemie, w którym problemy detekcji losowości sygnałów lub oceny zakłóceń wiążemy z charakterystykami zmienności lub gładkości. W ramach badanego przykładu opracowana metoda pozwala na efektywną identyfikację komponentów destrukcyjnych dla wyników prognozy. Prezentowane metody otwierają duże możliwości dalszych badań teoretycznych oraz aplikacyjnych w tym obszarze.

## BIBLIOGRAFIA

- Breiman L. (1996) Bagging Predictors. *Machine Learning*, 24, 123-140.  
Clements R. T. (1989) Combining Forecasts: A Review and Annotated Bibliography. *International Journal of Forecasting*, 5, 559-581.

- Comon P., Jutten C. (2010) Handbook of Blind Source Separation: Independent Component Analysis and Applications. Academic Press.
- Gatnar E. (2008) Podejście wielomodelowe w zagadnieniach dyskryminacji i regresji. PWN.
- Hamilton J. D. (1994) Time Series Analysis. Princeton University Press, Princeton.
- Hurst H. E. (1951) Long Term Storage Capacity of Reservoirs. Transactions of the American Society of Civil Engineers, 116, 770-799.
- Hyvarinen A., Karhunen J., Oja E. (2001) Independent Component Analysis. John Wiley, New York.
- Oja E. (1995) PCA, ICA and Nonlinear Hebbian Learning. Proc. Int. Conf. on Artificial Neural Networks ICANN-95, Oct. 9-13, France, 89-94.
- Shiryaev A. N. (1999) Essentials of Stochastic Finance: Facts, Models, Theory. Singapore, World Scientific.
- Szupiluk R., Wojewnik P., Ząbkowski T. (2004) Model Improvement by the Statistical Decomposition. Lecture Notes in Computer Science, 3070, 1199-1204.
- Szupiluk R. (2013) Dekompozycje wielowymiarowe w agregacji predykcyjnych modeli data mining. Oficyna Wydawnicza Szkoły Głównej Handlowej w Warszawie.
- Szupiluk R., Wojewnik P., Ząbkowski T. (2010) Noise Detection for Ensemble Methods. Lecture Notes in Artificial Intelligence, 6113, 471-478.
- Therrien C. W. (1992) Discrete Random Signals and Statistical Signal Processing. Prentice Hall, New Jersey.
- Vasegi S. V. (1997) Advanced Signal Processing and Digital Noise Reduction. John Wiley and Sons, Chichester.

#### IDENTIFICATION OF DESTRUCTIVE COMPONENTS IN PREDICTIVE MODELS WITH A MULTI-MODEL APPROACH

**Abstract:** In this paper we present a method of identification of destructive components in predictive models. This method may be applied in case of a multi-model approach and uses algorithms of blind signal separation. The evaluation of the characteristics of individual components will be based on the proposed metrics for evaluating the variation or smoothness of signals. In order to confirm the effectiveness of the presented method, we will present a practical experiment in which the results of the forecast of short-term electricity consumption are improved.

**Keywords:** prediction, blind separation, ensemble methods, Theta noise measure

## DYNAMICZNA ANALIZA SHIFT SHARE – OCENA ZMIAN STRUKTURY PRODUKCJI ZWIERZĘCEJ W POLSCE

**Agnieszka Tłuczak (ORCID: 0000-0001-6217-8822)**

Wydział Ekonomiczny  
Uniwersytet Opolski w Opolu  
e-mail: atluczak@uni.opole.pl

**Streszczenie:** Dynamiczna metoda przesunięć udziałów uwzględnia zmienność wag w kolejnych okresach. Metoda ta została zaproponowana w 1988 przez Barffa i Knighta. Ideą metody jest przeprowadzenie w sposób rekurencyjny obliczeń dla każdego z dwóch kolejnych okresów, aby ostatecznie dokonać agregacji wyników cząstkowych. Dodatkowo metoda ta pozwala na określenie tendencji w zakresie przemian gospodarki regionalnej. Celem artykułu jest przedstawienie analizy składników zmiany i możliwości jej wykorzystania do diagnostyki potencjału produkcyjnego rolnictwa regionu, a w szczególności wskazanie obiektów wyspecjalizowanych oraz charakteru zmian strukturalnych w zakresie badanego zjawiska.

**Słowa kluczowe:** konkurencyjność, rolnictwo, wagi Barffa-Knighta

### WSTĘP

Produkcja zwierzęca jest procesem przetwórczym produktów roślinnych na produkty zwierzęce. Obejmuje chów bydła, trzody chlewnej, owiec, drobiu, koni i innych gatunków zwierząt gospodarskich w celu uzyskania żywca rzeźnego oraz innych produktów zwierzęcych, np. mleka, jaj, wełny. W ujęciu globalnym jest ona sumą produkcji uzyskanej we wszystkich gospodarstwach rolnych, tj. stanowiących własność sektora prywatnego i publicznego [Tłuczak 2015, *Metodyka...* 2008]. Udział towarowej produkcji zwierzęcej w towarowej produkcji ogółem w 2015 roku zmniejszył się, było to wynikiem spadku udziału towarowej produkcji mleka (o 2,5 pkt. proc.) oraz znacznie mniejszego spadku udziału żywca wieprzowego (o 0,4 pkt. proc.) i jaj kurzych (o 0,2 pkt. proc.). Wartość produkcji globalnej sektora rolnego w cenach bieżących w 2015 roku, w porównaniu z rokiem poprzednim, zmniejszyła się o 6,7%. Zmniejszenie

wartości produkcji globalnej było wynikiem spadku wartości produkcji roślinnej o 7,4%, jak również produkcji zwierzęcej o 6,0%. Obniżenie wartości produkcji zwierzęcej wynikało przede wszystkim z głębokiego spadku wartości: przyrostu stada podstawowego i obrotowego – o ponad 80%, produkcji żywca wieprzowego – 6,8%, żywca cielęcego – o blisko 34%. W porównaniu z rokiem poprzednim, na skutek wzrostu podaży żywca i dekoniunktury na rynku światowym obniżyły się również ceny żywca wieprzowego. Relacje cen skupu żywca wieprzowego do cen zbóż nadal kształtowały się poniżej poziomu opłacalności. W warunkach dalszego rozwoju produkcji drobiarskiej i utrzymującej się wysokiej krajowej podaży żywca drobiowego, ceny drobiu rzeźnego kształtowały się również na poziomie niższym niż w 2014 r. Na rynku wołowiny odnotowano natomiast niewielki wzrost cen. Ponad 3-procentowy wzrost produkcji zwierzęcej wynikał ze zwiększenia produkcji żywca rzeźnego ogółem – o 5,1%, na który złożył się wzrost produkcji żywca drobiowego (o 6,4%), wieprzowego (o 1,9%), wołowego łącznie z cielęcym (o 10,9%) [Rolnictwo w 2015 roku 2016].

Zasadniczym celem artykułu jest analiza zmian struktury produkcji zwierzęcej w województwach Polski w latach 2010-2015, według gatunków żywca (wołowy, drobiowy, wieprzowy) z zastosowaniem klasycznej oraz dynamicznej analizy przesunięć udziałów (shift-share analysis). W opracowaniu dokonano oceny tempa wzrostu wielkości zjawiska. Ponadto dokonano analizy składników zmiany i możliwości jej wykorzystania do diagnostyki potencjału produkcyjnego rolnictwa regionu, zwrócono również uwagę na możliwość wskazania obiektów wyspecjalizowanych oraz charakteru zmian strukturalnych w zakresie badanego zjawiska.

## METODA BADAWCZA

Metody i modele analizy przesunięć udziałów (Shift-Share Analysis, SSA) należą do grupy analiz strukturalno-geograficznych [Suchecki 2010, Szewczyk, Łobos 2011, Tłuczak 2015]. Analizy strukturalno – geograficzne pozwalają na badanie i ocenę poziomu rozwoju danego regionu na tle poziomu rozwoju obszaru referencyjnego (np. kraju). Klasyczna metoda analizy przesunięć udziałów wprowadzona została do literatury przez Dunna [Dunn 1960] oraz Perloff, Dunna, Lamparda i Mutha [Perloff i inni 1960]. Pozwala ona na zrozumienie różnic w stopie wzrostu ekonomicznego między regionami [Antczak, Lewandowska-Gwarda 2015].

W klasycznej analizie przesunięć udziałów bada się kształtowanie zmiennej TX skwantyfikowanej w postaci złożonej przyrostu bezwzględny lub tempa zmian [Otsuko 2016]. Stosowanie w badaniach analizy SSA polega na dekompozycji całkowitej zmiany zlokalizowanej zmiennej na trzy części składowe odzwierciedlające [Trzpiot i in. 2013, Suchecki 2010, Tłuczak 2015]:

- część krajową (globalną) rozwoju regionalnego  $M_{ri}$ ,
- część strukturalną (zmiany struktury sektorowej)  $E_{ri}$ ,



- część lokalną rozwoju regionalnego  $U_{ri}$ .

Identyfikacja przewagi komparatywnej sektorów i usług staje się kluczowa a rola tej przewagi jest tym większa im słabiej rozwinięty jest region.

Klasyczne równanie shift share dla przyrostów rozważanej zmiennej przyjmuje postać [Esteban-Marquillas 1972, Silajdzic, Hashi 2011]:

$$\Delta x_{ri} = M_{ri} + E_{ri} + U_{ri}, \quad (1)$$

jednocześnie:

$$x_{ri}^* - x_{ri} = x_{ri} tx_{..} + x_{ri} (tx_{ri} - tx_{..}) + x_{ri} (tx_{ri} - tx_{i}), \quad (2)$$

gdzie:

$x_{ri}^*$  - obserwacje zmiennej X w r-tym regionie i i-tym sektorze w okresie końcowym,

$x_{ri}$  - obserwacje zmiennej X w r-tym regionie i i-tym sektorze w okresie początkowym,

$tx_{ri} = \frac{x_{ri}^* - x_{ri}}{x_{ri}}$  - indywidualne tempo wzrostu zmiennej X w i-tym sektorze

i r-tym regionie,

$tx_{i} = \frac{\sum_{r=1}^R (x_{ri}^* - x_{ri})}{\sum_{r=1}^R x_{ri}}$  - przeciętne tempo wzrostu zmiennej X w i-tym

sektorze,

$tx_{..} = \frac{\sum_{r=1}^R \sum_{i=1}^S (x_{ri}^* - x_{ri})}{\sum_{i=1}^S \sum_{r=1}^R x_{ri}}$  - przeciętne tempo wzrostu zmiennej X w kraju

w danym okresie.

Wersja dynamiczna analizy shift share została zaproponowana po raz pierwszy w 1988 roku przez Barffa i Knighta [Barff, Knight 1988]. Barff i Knight III zastosowali dynamiczną analizę shift, aby zmierzyć wzrost zatrudnienia w Nowej Anglii w latach 1939-1984. Wyniki ich badań wskazały, że głównymi czynnikami wpływającymi na wzrost zatrudnienia były znaczne różnice między regionalnymi i krajowymi stopami wzrostu.

Koncepcja metody zakłada zmienność wag w kolejnych okresach analizy, zaś obliczenia przeprowadza się w sposób rekurencyjny, biorąc każdorazowo dwa kolejne lata w kilkuletnim badaniu. Następnie dokonuje się agregacji wyników cząstkowych:

$$\sum_k \Delta x_{ri} = \sum_k M_{ri} + \sum_k E_{ri} + \sum_k U_{ri}. \quad (3)$$

Zaproponowane przez Barffa i Knighta podejście pozwala na otrzymanie lepszych, bardziej wiarygodnych rezultatów ze względu na alokację wzrostu między składowymi. Dodatkowo podejście to dostarcza dodatkowych informacji nt. przemian gospodarki regionalnej oraz pozwala na określenie jej tendencji rozwojowej [Suchecki 2010].

## WYNIKI BADAŃ

Analiza struktury produkcji zwierzęcej została wykonana dla wszystkich województw w odniesieniu do poziomu rozwoju tego zjawiska w kraju. Dane do badania pobrano z Banku Danych Lokalnych Głównego Urzędu Statystycznego. Zmienną badaną oraz referencyjną jest produkcja zwierzęca według 4 gatunków mięsa w latach 2010-2015 dla województw w Polsce. Obliczenia przeprowadzono przy wykorzystaniu obu prezentowanych metod.

Tabela 1 prezentuje wyniki klasycznej i dynamicznej analizy shift share. Porównując tempa wzrostu/spadku regionalnego poszczególnych województw z przeciętnym, krajowym wzrostem (18,7%) w analizowanych latach 2010 i 2015 można zaobserwować, że we wszystkich województwach odnotowano wyższy wzrost produkcji zwierzęcej od tempa krajowego.

Tabela 1. Wyniki klasycznej i dynamicznej SSA oraz różnice przyrostów wielkości produkcji zwierzęcej między wynikami obu analiz według województw i efektów

Województwo \ SSA	Klasyczna SSA			Dynamiczna SSA			Różnice przyrostów		
	M	E	U	$\Delta M$	$\Delta E$	$\Delta U$	M- $\Delta M$	E- $\Delta E$	U- $\Delta U$
Dolnośląskie	24,260	-34,460	-42,177	21,828	-32,028	-41,870	2,432	-2,432	-0,306
Kujawsko-pomorskie	78,822	-18,922	3,023	82,791	-22,891	-5,225	-3,969	3,969	8,248
Lubelskie	58,145	-46,645	-31,694	53,083	-41,583	-24,585	5,062	-5,062	-7,109
Lubuskie	25,986	-4,586	-21,187	26,628	-5,228	-19,285	-0,641	0,641	-1,902
Łódzkie	78,916	45,184	58,198	83,787	40,313	50,729	-4,872	4,872	7,469
Małopolskie	32,459	-72,159	-66,103	23,534	-63,234	-54,637	8,925	-8,925	-11,466
Mazowieckie	134,565	259,335	237,088	160,302	233,598	203,396	-25,737	25,737	33,692
Opolskie	29,889	50,589	47,569	24,938	45,638	41,744	4,951	4,951	5,825
Podkarpackie	21,690	-41,190	-39,065	18,440	-37,940	-36,805	3,249	-3,249	-2,260
Podlaskie	53,473	-13,973	-11,054	54,290	-14,790	-12,743	-0,816	0,816	1,689
Pomorskie	52,742	44,558	57,400	56,236	41,064	57,368	-3,494	3,494	0,031
Śląskie	33,266	-7,466	-19,721	30,871	-5,071	-15,352	2,395	-2,395	-4,370
Świętokrzyskie	25,161	16,539	16,210	24,457	17,243	15,763	0,703	-0,703	0,447
Warmińsko-mazurskie	65,594	-11,894	-40,741	62,471	-8,771	-29,484	3,123	-3,123	-11,257
Wielkopolskie	211,417	-28,817	-1,597	204,413	-21,813	1,707	7,004	-7,004	-3,304
Zachodniopomorskie	41,315	-34,915	-51,010	39,630	-33,230	-47,233	1,685	-1,685	-3,777

M-część krajowa, E-część strukturalna, U-część lokalna.

Źródło: opracowanie własne na podstawie danych GUS

Największy wzrost produkcji zwierzęcej nastąpił w województwach mazowieckim i wielkopolskim (zmiany te wynikały przede wszystkim z dużego wzrostu produkcji trzody chlewnej). Województwa podkarpackie i dolnośląskie to natomiast województwa gdzie ten wzrost był również powyżej krajowego, lecz najmniejszy w stosunku do pozostałych województw. Zmiany te wynikały zarówno ze zmian struktury produkcji (efekty strukturalne), jak i ze zmian wewnętrznych sytuacji konkurencyjności danego obszaru (efekty geograficzne). Najbardziej zbliżone pod względem średniego tempa wzrostu do tempa krajowego jest województwo podkarpackie (21,69%).

Tabela 2. Różnice przyrostów wielkości produkcji zwierzęcej między wynikami analizy klasycznej i dynamicznej według województw, sektorów i efektów

	Bydło			Cielęta			Trzoda			Drób		
	M- ΔM	E- ΔE	U-ΔU	M- ΔM	E- ΔE	U-ΔU	M- ΔM	E- ΔE	U-ΔU	M- ΔM	E- ΔE	U-ΔU
Dolnośląskie	0,78	-0,34	-0,85	0,04	-0,01	0,03	3,76	-4,21	0,44	-3,24	2,12	0,11
Kujawsko-pomorskie	-1,29	1,78	3,06	0,23	-0,16	0,45	0,42	-0,56	5,71	-4,71	3,27	-0,97
Lubelskie	2,09	-2,54	-3,13	1,01	-0,97	1,92	1,97	-2,11	-1,35	0,07	-0,04	-4,53
Lubuskie	0,01	-0,04	-0,18	-0,04	0,02	-0,21	0,77	-0,86	0,27	-1,95	1,4	-1,78
Łódzkie	-3,01	3,21	4,1	-0,12	0,09	-1,63	2,86	-2,91	2,37	-5,21	4,63	2,64
Małopolskie	4,75	-5,21	-6,82	0,98	-0,82	-0,81	1,73	-1,92	0,77	1,71	-1,62	-4,61
Mazowieckie	-2,69	3,12	4,94	1,64	-1,24	1,36	10,23	-11,5	-4,11	-35,29	34,52	31,50
Opolskie	0,45	-0,62	-0,70	-0,06	0,02	-0,25	3,99	-4,26	-1,76	0,85	-0,53	-3,11
Podkarpackie	1,24	-1,39	-1,56	0,84	-0,65	1,56	-0,59	0,67	2,24	2,07	-1,95	-4,50
Podlaskie	2,69	-2,84	-1,31	0,92	-0,89	1,66	-0,12	0,17	0,13	-4,23	4,29	1,22
Pomorskie	-0,13	0,18	0,72	0,08	-0,03	-0,18	2,93	-2,87	-3,47	-5,18	6,33	2,98
Śląskie	1,29	-1,32	-1,73	0,02	0,00	-0,38	1,99	-1,52	-1,14	-0,74	0,8	-1,11
Świętokrzyskie	0,88	-0,98	-0,48	0,26	-0,11	-0,79	0,99	-1,06	0,46	-1,91	1,28	1,25
Warmińsko-mazurskie	0,12	-0,16	0,51	-0,19	0,14	-0,56	2,96	-3,24	-2,39	0,29	-0,21	-8,82
Wielkopolskie	-2,19	2,37	3,93	-0,21	0,31	-1,65	20,53	-21,6	3,54	-13,24	11,04	-9,12
Zachodniopomorskie	0,36	-0,49	-0,49	-0,01	0,04	-0,49	5,15	-6,07	-1,67	-4,21	3,79	-1,13

M-część krajowa, E-część strukturalna, U-część lokalna.

Źródło: opracowanie własne na podstawie danych GUS

Przykładem opisanym szczegółowo będzie województwo opolskie. Wzrost poziomu produkcji zwierzęcej (29,89%) w województwie opolskim, czyli o 11,13% ponad przeciętne tempo wzrostu krajowego, spowodowany był w dużym

stopniu zmianami strukturalnymi działalności restrukturyzacyjnej (50,59%), i w nieznacznie mniejszym stopniu przez zmiany wewnętrzne związane z konkurencyjnością z innymi regionami (47,57%). Można zauważyć, że zmiany w poziomie analizowanej zmiennej, w każdym z badanych województw, wywołane były w równym stopniu zmianami wynikającymi ze zmian wewnętrznych sytuacji konkurencyjności danego obszaru (efekt geograficzny) jak i zmianami wynikającymi ze zmian strukturalnych (efekt strukturalny). Wyniki dynamicznej analizy shift share są zbieżne z wynikami uzyskanymi za pomocą klasycznej SSA. Różnice w przypadku poszczególnych województw wahają się od 3% do 33%, najmniejsze różnice wystąpiły w przypadku województwa małopolskiego (średnia dla efektów to -3,82%), natomiast największe różnice dotyczą województwa mazowieckiego (średnia 11,23%).

Rozważając różnice przyrostów wielkości produkcji zwierzęcej między wynikami analizy klasycznej i dynamicznej według województw, sektorów i efektów widoczne są rozbieżności. Największe odchylenia między wynikami klasycznej i dynamicznej analizy shift share odnotowano w województwie mazowieckim, w sektorze drobiu i trzody chlewnej w odniesieniu do efektów strukturalnych i geograficznych. Najmniejsze różnice dotyczą województwa lubuskiego w sektorze bydła oraz województwa podlaskiego w sektorze trzody chlewnej (tabela 2).

W podejściu dynamicznym uzyskane wyniki pozwalają na uchwycenie zmian poziomu badanego zjawiska w całym okresie analizy (wagi z roku na rok). Natomiast wersja statyczna SSA uwzględnia jedynie zmiany wartości cechy z początku i końca okresu. Następuje pominięcie zmian zachodzących w poszczególnych latach (tabela 3). Średnie tempa wzrostu poziomu analizowanej zmiennej w klasycznej metodzie przesunięć udziałów są wyższe od temp metody dynamicznej shift share. Potwierdza to konieczność stosowania metody dynamicznej. Rezultaty klasycznej analizy SSA zależą od wyboru systemu wag, które mogą powodować obciążenie rezultatów. Można zauważyć, iż w niektórych przypadkach wartości różnic efektów sięgają od 0,02% (cielęta) do 12,12% (trzoda). Tempo krajowe w analizie dynamicznej różni się od tempa krajowego w analizie statycznej w ramach każdego gatunku mięsa. W przypadku efektów geograficznych (U), które nazywane są również wskaźnikami konkurencyjności lokalnej, można analizować istnienie konkurencyjności r-tego województwa pod względem zmian w i-tym wariancie zjawiska w relacji do województw sąsiadujących. W rozważanym województwie opolskim w przypadku produkcji trzody otrzymano wartości dodatnie jedynie w latach 2011/2010. Oznacza to, że sąsiednie obszary wywierają na negatywny wpływ na rozwój województwa opolskiego w zakresie produkcji trzody.

Sektory produkcji bydła, cieląt oraz drobiu charakteryzują się zmienną tendencją w zakresie zmian wartości efektu związanego z lokalnym rozwojem regionalnym. Największe dodatnie wartości omawianego efektu zaobserwowano w latach 2013/2012, tempo wzrostu produkcji żywca było wówczas o 8,77%

wyższe niż przeciętne tempo zmian na poziomie krajowym, a wynikało to z pozytywnych zmian.

Tabela 3. Wyniki dynamicznej SSA i porównanie z metodą klasyczną dla woj. opolskiego według sektorów i efektów składowych

DX					M				
Lata	Bydło	Cielęta	Trzoda	Drób	Lata	Bydło	Cielęta	Trzoda	Drób
2011/2010	-7,97	1,03	16,78	13,31	2011/2010	0,25	0,00	1,41	0,90
2012/2011	4,71	-0,91	-29,46	-36,90	2012/2011	-0,00	0,00	-0,01	-0,01
2013/2012	1,96	0,20	-23,08	16,97	2013/2012	-0,16	-0,00	-0,91	-0,57
2014/2013	-9,41	0,16	4,71	-20,34	2014/2013	1,81	0,03	8,08	6,94
2015/2014	-0,98	-0,42	-11,86	16,13	2015/2014	0,58	0,02	3,86	2,66
Suma	-11,69	0,06	-42,91	-10,83	Suma	2,49	0,05	12,44	9,94
2015/2010	-12,74	-0,03	-55,03	-3,47	2015/2010	2,94	0,03	16,43	10,47
Różnica	-1,05	-0,10	-12,12	7,36	Różnica	0,45	-0,02	3,99	0,52
E					U				
2011/2010	-4,15	0,49	7,28	6,69	2011/2010	-4,07	0,53	8,08	5,71
2012/2011	2,10	-0,50	-18,39	-14,99	2012/2011	2,61	-0,41	-11,06	-21,90
2013/2012	1,06	0,10	-11,08	8,77	2013/2012	1,06	0,10	-11,08	8,77
2014/2013	-5,61	0,06	-1,68	-13,64	2014/2013	-5,61	0,06	-1,68	-13,64
2015/2014	-0,78	-0,22	-7,86	6,73	2015/2014	-0,78	-0,22	-7,86	6,73
Suma	-7,39	-0,05	-31,74	-6,44	Suma	-6,79	0,06	-23,61	-14,33
2015/2010	-7,84	-0,03	-35,73	-6,97	2015/2010	-7,84	-0,03	-35,73	-6,97
Różnica	-0,45	0,02	-3,99	-0,52	Różnica	-1,05	-0,10	-12,12	7,36

M-część krajowa, E-część strukturalna, U-część lokalna.

Źródło: opracowanie własne na podstawie danych GUS

## PODSUMOWANIE

Analiza przesunięć udziałów jest metodą z pogranicza ekonometrii i statystyki, dokonującą podziału pewnej zmiany na różne czynniki. Posiada wielki potencjał rozwojowy pozwalający uwzględnić aspekty dynamiczne. W opracowaniu skoncentrowano się na wybranym aspekcie zmian struktury produkcji zwierzęcej. Właściwy kierunek produkcji rolnej przy zastanym potencjale regionu może przynieść producentom rolnym wysokie dochody. Zmiana kierunku produkcji, częstokroć kosztowna, może uchronić od poniesienia dodatkowych kosztów. Podjęta analiza produkcji zwierzęcej ma charakter niewyczerpujący. Każdy analizowany region ma swoje specyficzne cechy, które mają wpływ na tempo i kierunki jego rozwoju gospodarczego i społecznego. Zastosowane metody SSA, klasyczna i dynamiczna, mogą służyć jako narzędzia do oceny sytuacji ekonomicznej, czy gospodarczej regionów.

Uzyskane dwiema metodami wyniki różnią się nieznacznie od siebie. Efekty uzyskane klasyczną metodą przesunięć udziałów identyfikują efekty wpływające na zróżnicowanie województw Polski porównując ze sobą dwa skrajne okresy

(2010 i 2015 rok). Poprzez zastosowanie dynamicznej metody przesunięć udziałów możliwe jest zaobserwowanie tendencji zmian w zakresie poszczególnych efektów. Zaobserwowano, iż zróżnicowanie tempa zmian badanego zjawiska w województwach wiąże się zarówno z pozycją konkurencyjną danego regionu i jednocześnie ze zmian w strukturze produkcji zwierzęcej.

## BIBLIOGRAFIA

- Antczak E., Lewandowska-Gwarda K. (2015) Analysis of Emigration in Europe Using the Spatial Dynamic Shift-Share Method. *Folia Oeconomica Stetinensia*, 15(2), 7-26.
- Barff R. A., Prentice, L. K. (1988) Dynamic shift-share analysis. *Growth Chang*, 19, 1-10.
- Dunn E. S. (1960) A Statistical and Analytical Technique for Regional Analysis. *Papers and Proceedings of the Regional Science Association*, 6, 98-112.
- Esteban-Marquillas J. M. (1972) A Reinterpretation of Shift-Share Analysis. *Regional and Urban Economics*, 2(3), 249-255.
- Metodyka szacowania fizycznych rozmiarów produkcji zwierzęcej (2008) Główny Urząd Statystyczny, Warszawa.
- Otsuko A. (2016) Regional Energy Demand in Japan: Dynamic Shift-Share Analysis. *Energy, Sustainability and Society*, 6(1), 1-10.
- Perloff H. S., Dunn E. S., Lampard E. E., Muth R. F. (1960) *Regions, Resources and Economic Growth*. Johns Hopkins Press, Baltimore.
- Rolnictwo w 2015 roku. (2016) Główny Urząd Statystyczny, Departament Rolnictwa, Warszawa.
- Silajdzic S., Hashi I. (2011) Competitiveness of Transition Economies: Evidence from the Dynamic Shift-Share Analysis, *Challenges of Europe. International Conference Proceedings*, 657-712.
- Suchecky B. (red.) (2010) *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*. C. H. Beck, Warszawa, 162-194.
- Tłuczak, A. (2015) Przestrzenna metoda przesunięć udziałów w ocenie zróżnicowania produkcji zwierzęcej w Polsce. *Metody Ilościowe w Badaniach Ekonomicznych*, XVI/4, 171-180.
- Trzpiot G., Ojrzyńska A., Szoltysek J., Twaróg S. (2013) Wykorzystanie shift share analysis w opisie zmian struktury honorowych dawców krwi w Polsce. *Wielowymiarowe modelowanie i analiza ryzyka*. UE Katowice, 84-98.

**DYNAMIC SHIFT-SHARE ANALYSIS - ASSESSMENT  
OF STRUCTURE CHANGES IN ANIMAL PRODUCTION  
IN POLAND**

**Abstract:** The dynamic shift analysis takes into account the weight variability in two consecutive periods. This method was proposed in 1988 by Barff and Knight. The idea of this method is to recursively calculate for each of the two consecutive periods to finally aggregate the partial results. In addition, this method allows to identify trends in regional economic transformations. The aim of the article is to analyse the components of change and the possibility of its use for the diagnosis of the productive potential of agriculture in the region, and in particular the analysis allowed to identify specialized and nature of the structural changes within the analysed phenomenon.

**Keywords:** Barff-Knight wages, competitiveness, agricultural

## METODY PREZENTACJI DANYCH W SYSTEMACH BI

**Krzysztof Zmarzłowski**

Wydział Zastosowań Informatyki i Matematyki  
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie  
e-mail: krzysztof\_zmarzowski@sggw.pl

**Streszczenie:** W dobie XXI wieku mamy do czynienia z coraz większą ilością informacji wymagających odpowiedniego przetworzenia. Największym problemem jest kwestia selekcjonowania danych i wyodrębniania z nich tych najbardziej istotnych. Autorzy niniejszego opracowania podjęli próbę przedstawienia najważniejszych zasad prezentacji danych w nowoczesnych systemach Business Intelligence. Zasady te powinny pozwolić na ułatwienie czytelności zebranych informacji oraz ułatwienie ich interpretacji.

**Słowa kluczowe:** Raporty BI, prezentacja danych, wizualizacja dashboardów

### WSTĘP

Najbardziej problematycznym elementem we wdrożeniach systemów klasy Business Intelligence (BI) jest definicja i analiza potrzeb odbiorcy końcowego. Błędne decyzje podjęte na etapie projektowania, mogą obniżyć zarówno wartość takiego systemu jak i jego funkcjonalność. Istotą wdrożenia systemu jest więc przede wszystkim dokładna analiza ewentualnego sposobu wykorzystania danych przez użytkownika końcowego.

Głównym celem tego artykułu jest przybliżenie obecnie dostępnych metod prezentacji danych w raportach systemów klasy BI. Dodatkowo zaprezentowano przykładowe zastosowania oraz podstawowe praktyki dobrego projektowania raportów, wraz z tłem psychologicznym na temat percepcji wzrokowej człowieka.

### TŁO HISTORYCZNE RAPORTÓW BI/DASHBOARDÓW

Nawiązując do słownika języka angielskiego, słowo „dashboard” w kontekście systemów Business Intelligence określa interfejs graficzny bądź stronę internetową, która wyświetla użytkownikowi dane, zazwyczaj w prys-



tępną, łatwą do przeczytania formę. Na dzień dzisiejszy słowo to nie ma poprawnego odpowiednika w języku polskim, a jego dostępne tłumaczenia brzmią: „deska rozdzielcza” lub „tablica wskaźników” [www.en.oxforddictionaries.com].

Za pierwsze systemy używające dashboardów uważa się EISy (Executive Information Systems), które pojawiły się na rynku wraz z początkiem lat osiemdziesiątych. Celem EISów było dostarczenie użytkownikowi danych wysokiego poziomu, takich jak roczna wartość sprzedaży danego produktu względem planu lub procent przesyłek dostarczonych na czas w ciągu roku. Odbiorcami tego typu systemów była zazwyczaj bardzo wąska grupa kierownictwa firmy [Watson 1993]. Na rysunku 1 przedstawiono przykład raportu z tego rodzaju systemów.

Rysunek 1. Rekonstrukcja przykładowego dashboardu systemu klasy EIS firmy Pilot Software

1990 Scorecard					
ADI					
Qtr 1 1990					
Line Item		Actual	Budget	Variance	% Var.
<b>FINANCIALS:</b>					
Sales	\$M	39.40	41.00	-1.60	-3.90 #
Sales Growth (YTY)	%	-5.40	-1.60	-3.80	-237.50 #
Contribution Margin	%	10.40	11.60	-1.20	-10.34 #
ROA (Contribution Margin)	%	9.70	11.20	-1.50	-13.39 #

Źródło: [Schneiderman 1990]

Głównymi problemami EISów pod koniec lat osiemdziesiątych była niska wydajność, brak przystępnych narzędzi do badania jakości danych oraz wysoki poziom wiedzy programistycznej, wymagany do tworzenia bądź edycji raportów. Raporty te składały się z głównie z tabel, wykresów słupkowych, wykresów liniowych i były statyczne, tzn. nie dawały np. możliwości filtrowania danych. Zamyśl twórców EISów był słuszny i mocno wyprzedzał swoją epokę, jednak pod koniec lat osiemdziesiątych, technologie bazodanowe oraz wydajność nawet najlepszych komputerów klasy mainframe była zdecydowanie niewystarczająca do wydajnej pracy. Z tych powodów rozwój EISów w latach osiemdziesiątych został wstrzymany i częściowo powrócił dopiero w latach dziewięćdziesiątych, wraz z rozwojem baz i hurtowni danych, tzw. Data Warehouse.

Lata dziewięćdziesiąte to czas dynamicznego rozwoju technologii Data Warehouse, kostek OLAP (online analytical processing) oraz utworzenia terminu Business Intelligence, który do dzisiaj nie ma jednoznacznej definicji. Jedno z wielu znaczeń tego terminu na dzień dzisiejszy brzmi:

- Przekształcanie danych w informacje, a informacji w wiedzę w celu optymalizacji działania procesów biznesowych i całej organizacji.

- Zbiór praktyk, metodyk, narzędzi i technologii informatycznych, służących zbieraniu i integrowaniu danych w celu dostarczania informacji i wiedzy właściwym osobom, we właściwym miejscu oraz we właściwym czasie [www.bi.pl].
- Proces przekształcania danych w informacje, a informacji w wiedzę, która może być wykorzystana do zwiększenia konkurencyjności przedsiębiorstwa [www.codecdss.co.uk].

Oprócz rozwoju technologicznego, wiele firm zaczęło przywiązywać więcej wagi do tego, jak pracować z danymi, aby w możliwie najprzystępniejszy sposób obrazować kondycję firmy oraz móc wyznaczać jej cele. W wyniku tej potrzeby Kaplan oraz Norton zbudowali metodologię tzw. karty wyników (ang. Balanced Scorecard). Jej podstawą były kluczowe wskaźniki efektywności z ang. Key Performance Indicators (KPI).

Idea Balanced Scorecard'u polegała na prezentacji danych finansowych oraz operacyjnych w formie jednego zbiorczego raportu, czyli dashboard'u. Powodem integracji tak różnorodnych danych w jednym raporcie, było spostrzeżenie autorów, że zarząd firmy nie opiera swoich decyzji tylko i wyłącznie na danych finansowych. Sama informacja o sprzedaży danego produktu nie jest w stanie w pełni odzwierciedlić rzeczywistości i nie daje możliwości wyciągnięcia poprawnych wniosków. Aby raport Balanced Scorecard odzwierciedlał rzeczywistość jak najlepiej, Kaplan i Norton spędzili rok, współpracując z dwunastoma firmami i opracowując zbiór uniwersalnych metryk, które mogłyby być w stanie zobrażować kondycję każdej firmy. Rezultatem tych badań były cztery podstawowe pytania, na które powinien odpowiadać Balanced Scorecard:

1. Jak widzą nas klienci? (perspektywa klienta).
  2. W czym jesteśmy najlepsi? (perspektywa wewnętrzna).
  3. Czy jesteśmy w stanie ulepszać firmę i tworzyć "wartość dodaną"? (innowacje i perspektywa rozwoju).
  4. Jak wyglądamy z perspektywy akcjonariuszy? (perspektywa finansowa).
- [Kaplan, Norton 1991], [Shivakumar 2014].

Tworząc metryki które odpowiadają na tak postawione pytania, firma jest w stanie zawrzeć niemal pełen obraz postępu i rozwoju na przestrzeni czasu. Ponadto, zgrupowanie metryk w cztery kategorie sprawia, iż dane stają się dużo czytelniejsze dla użytkownika końcowego.

Mimo pierwszych rozwiniętych metodologii pracy z danymi, takich jak Balanced Scorecard i pierwszych przystępnych narzędzi w postaci kostek OLAP, dashboard'y nie spopularyzowały się mocno w latach dziewięćdziesiątych. Nastąpiło to dopiero po roku 2001 za sprawą skandalu firmy Enron [www.pl.wikipedia.org/wiki/Enron].

Rezultatem wyników finansowych uzyskanych przez Enron było zwrócenie uwagi wielu firm na sposoby kontroli stanu finansów firmy, przy użyciu systemów BI. Korporacje z branży IT takie jak IBM czy SAP umiejętnie wyczuły nastroje

rynku i zaczęły dynamicznie inwestować w rozwój takich narzędzi oraz ich marketing. Pomijając oczywisty rozwój technologiczny branży IT, filozofia dashboardów, z których korzystamy obecnie, nie zmieniła się od roku 2001. Warto jednak odnotować, iż w międzyczasie powstawała inna koncepcja pracy z danymi niż klasyczny raport przygotowywany przez zespół programistów. Rozwiązanie to dawało użytkownikowi końcowemu możliwość wielowymiarowych oraz dynamicznych analiz.

Jednym z takich narzędzi okazała się tabela przestawna, która swoje początki ma w roku 1986. To programista Pito Salas z firmy Lotus zaczął pracować nad koncepcją narzędzia, które rozdzielało pracę z danymi na trzy obszary:

1. dane źródłowe wprowadzane przez użytkownika,
2. formuły kalkulacji danych użytkownika,
3. finalny rezultat wybranych miar.

Rezultatem prac była pierwsza implementacja tabeli przestawnej w arkuszu kalkulacyjnym, zaprezentowana jako funkcjonalność oprogramowania Lotus Improv w roku 1991 [Michael, Jelen 2005].

Rysunek 2. Tabela przestawna w Lotus Improv

Region	Year	1989	1989	1990	1991
Galaxy	Vending	10.01%	8.85%	8.72%	8.84%
	Grocery	4.15%	3.94%	4.28%	4.74%
	Supermarket	0.00%	15.76%	15.87%	15.95%
	All channels	14.16%	28.51%	28.84%	29.53%
Snackers	Vending	10.01%	8.85%	8.72%	8.84%
	Grocery	3.86%	3.48%	3.69%	3.88%
	Supermarket	20.60%	16.47%	15.84%	14.87%
	All channels	34.46%	28.84%	28.84%	27.58%
Mintz	Vending	8.86%	8.96%	6.70%	6.47%
	Grocery	2.80%	2.55%	2.87%	2.89%
	Supermarket	16.60%	13.99%	13.41%	13.15%
	All channels	28.18%	23.20%	22.88%	22.29%
Paydirt	Vending	9.73%	8.10%	5.38%	6.60%
	Grocery	3.16%	3.02%	3.35%	3.66%
	Supermarket	14.31%	11.83%	11.62%	11.42%
	All channels	23.18%	18.95%	20.34%	20.69%
All products	Vending	34.33%	29.23%	29.80%	29.74%
	Grocery	14.16%	12.39%	13.87%	14.87%
	Supermarket	51.80%	57.77%	56.94%	55.38%
	All channels	100.00%	100.00%	100.00%	100.00%

Źródło: opracowanie własne na podstawie Wikipedia, Lotus Improv

Dla systemów BI tabela przestawna zaczęła odgrywać istotną rolę w połowie lat 90. Wtedy to producenci oprogramowania zaczęli tworzyć dodatki do arkuszy kalkulacyjnych, które umożliwiały łączenie się z kostkami OLAP. Przy tego typu architekturze, wielowymiarowe kostki analityczne były odpowiedzialne za:

1. odpytywanie baz danych w celu przygotowania danych niezbędnych do kalkulacji poszczególnych miar,
2. przetrzymywanie formuł kalkulacji oraz przeprowadzanie obliczeń na podstawie danych.

Dzięki takiej architekturze, użytkownik końcowy zawsze dostaje gotowy produkt w postaci tabeli przestawnej z przedefiniowanymi wymiarami i miarami, bez potrzeby wykonywania punktów 1 i 2 ręcznie w arkuszu kalkulacyjnym [Jelen 2005].

Wśród producentów oprogramowania Business Intelligence obecnie najnowszym trendem jest koncepcja łączenia atutów elastyczności pracy z danymi, jaką oferują tabele przestawne oraz możliwość prezentacji graficznej, jaką oferują dashboards. Istotną zmianą jest również sposób postrzegania użytkowników końcowych.

W klasycznych systemach raportowych poprzedniej generacji, role użytkowników były podzielone na dwie mocno rozgraniczone sfery:

1. wysokiej klasy specjalistów IT odpowiedzialnych za utrzymanie systemu oraz tworzenie raportów,
2. odbiorców - pracowników, analityków oraz menedżerów.

W tzw. systemach self service z uwagi na niższy poziom wymaganej wiedzy technicznej, głównym twórcą raportów powinien być analityk, bardziej rozumiejący potrzeby biznesu niż wyspecjalizowany programista. Odbiorcami raportów powinni być nadal pracownicy i zarząd, podczas, gdy ilość wyspecjalizowanych pracowników z sektora IT powinna być ograniczona do niezbędnego minimum, aby prawidłowo utrzymać system i zapewnić pomoc ekspercką analitykom.

Oprócz ulepszeń istniejących komponentów do przetwarzania oraz wizualizacji danych znanych z systemów raportowych poprzednich generacji, obecnie producenci wprowadzają wiele interesujących funkcjonalności, takich jak:

- możliwość komentowania oraz przysyłania raportów (wzorowanie się na mediach społecznościowych),
- wyszukiwarki semantyczne,
- wykresy umożliwiające “drążenie” w danych (ang. - drill down charts),
- obsługa narzędzi do zaawansowanych analiz statystycznych (np. języka R),
- aplikacje mobilne i responsywne raporty, umożliwiające przeglądanie raportów na smartfonach oraz wiele innych.

## METODY PREZENTACJI DANYCH

Zgodnie z chińską sentencją jeden obraz jest wart tysiąc słów, jedną z najlepszych metod pozwalającą zrozumieć dane jest ich wizualizacja. Liczby w postaci obrazu, nie są jednak dla każdego odbiorcy metodą najprzystępniejszą.

Stephen Few, jeden z liderów w dziedzinie wizualizacji danych, wyodrębnia trzy obszary które wpływają na dobór odpowiedniej metody prezentacji danych do danego zastosowania:

1. Cel raportowania:

- wyszukiwanie (ang. Lookup) - aby umożliwić odnalezienie poszukiwanych faktów,
  - narracja - aby zakomunikować zbiór powiązanych ze sobą faktów,
  - monitoring - aby zaktualizować świadomość odbiorcy na dany temat,
  - analityczny kierowany - aby przeprowadzić analizę danych przy użyciu narzędzi zaprojektowanych do udzielania odpowiedzi na wybrane pytania,
  - analityczny badawczy - aby zbadać zbiór danych w poszukiwaniu użytecznych informacji, niekoniecznie odpowiadając na postawione pytania,
  - analiza predykcyjna - aby szacować prawdopodobne wartości na podstawie określonych warunków.
2. Platforma wizualizacyjna:
- raport do wyszukiwania (ang. Lookup Report),
  - raport objaśniający (ang. Explanatory Report) - do celów narracyjnych,
  - infografika - do celów narracyjnych,
  - raport do prezentacji na żywo - do celów narracyjnych,
  - dashboard - do monitorowania,
  - aplikacja analityczna - do celów analitycznych (kierowanych),
  - aplikacja do eksploracji danych - do celów analitycznych (badawczych),
  - model predykcyjny - do celów analizy predykcyjnej.
3. Cechy raportu:
- poziom interakcji w skali od 1-5 (1-brak interakcji, 5-pełna elastyczność dla użytkownika końcowego),
  - częstotliwość użycia w skali od 1-3 (1-bardzo rzadko, 3-regularnie),
  - częstotliwość aktualizacji danych w skali od 1-4 (1-nigdy, 4-w czasie rzeczywistym bądź prawie w czasie rzeczywistym),
  - czas reakcji (jak szybko odbiorca raportu musi podjąć działania na podstawie tego raportu) w skali od 1-2 (1-działania mogą zostać podjęte później, 2-wymagana jest natychmiastowa reakcja),
  - poziom umiejętności analitycznych odbiorców raportów w skali od 1-3 (1-niskie, 3-wysokie),
  - środek przekazu: 1-5 (1-druk, 2-duży ekran (projektor), 3-kran komputera, 4-tablet, 5-smartfon) [Few 2004, 2013, 2014].

## JAK TWORZYĆ DOBRE RAPORTY

Wszystkie cechy dobrze zaprojektowanego raportu mają swoje źródło w ogólnych zasadach dobrego projektowania, uniwersalnych dla wielu gałęzi przemysłu, począwszy od architektury przez przedmioty użytku codziennego, aż po interfejsy aplikacji komputerowych. Na początku lat siedemdziesiątych, Dieter

Rams - jeden z najbardziej uznanych projektantów wzornictwa przemysłowego, sformułował następujące dziesięć cech dobrego wzornictwa, które stały się inspiracją dla wielu innych dziedzin projektowych:

1. „Dobre wzornictwo jest innowacyjne”- możliwości innowacji nie są pod żadnym pozorem wyczerpane. Rozwój technologiczny zawsze oferuje nowe możliwości do wykorzystania w innowacyjnym wzornictwie. Innowacyjne wzornictwo jednak zawsze rozwija się w parze z innowacyjną technologią i nigdy nie jest celem samym w sobie.
2. „Dobre wzornictwo czyni produkt użytecznym”- produkt kupuje się po to, by go używać. Musi on spełnić określone kryteria, nie tylko funkcjonalne, ale także psychologiczne i estetyczne. Dobre wzornictwo podkreśla użyteczność produktu bez względu na ewentualne czynniki, które mogłyby odwracać od tego uwagę.
3. „Dobre wzornictwo jest estetyczne” - jakość estetyczna produktu jest integralna z jego użytecznością, ponieważ produkty, których używamy, codziennie wpływają na naszą osobowość i samopoczucie. Tylko dobrze wytworzone produkty mogą być piękne.
4. „Dobre wzornictwo czyni produkt zrozumiałym” - wyjaśnia on strukturę produktu. A więc klient widząc produkt natychmiast domyśla się do czego dany produkt może być użyty.
5. „Dobre wzornictwo jest skromne” - produkty, które spełniają swoją funkcję, są jak narzędzia. Nie są one ani obiektami dekoracyjnymi, ani dziełami sztuki. Powinny być zatem zaprojektowane w sposób neutralny i powściągliwy, aby zostawić użytkownikowi przestrzeń do interpretacji.
6. „Dobre wzornictwo jest szczerze” - dobre wzornictwo nie udaje, że produkt jest bardziej innowacyjny, wartościowy, niż jest w rzeczywistości. W dobrym wzornictwie nie próbuje się manipulować konsumentem poprzez obietnice, które nie mogą zostać spełnione.
7. „Dobre wzornictwo jest długowieczne” - unika bycia modnym, a przez to nigdy nie staje się eksponatem. W przeciwieństwie do modnego wzornictwa, może przetrwać wiele lat – nawet w dzisiejszym konsumpcyjnym świecie.
8. „Dobre wzornictwo jest dokładne, do najdrobniejszych szczegółów” - nic nie może być dowolne czy pozostawione przypadkowi. Dokładność i precyzja w procesie projektowania dowodzą szacunku dla klienta.
9. „Dobre wzornictwo jest przyjazne środowisku” - wzornictwo odgrywa znaczącą rolę w dbaniu o środowisko naturalne. Może oszczędzać zasoby, redukować fizyczne zanieczyszczenie oraz wizualne skażenie naszego środowiska podczas całego cyklu życia produktu.
10. „Dobre wzornictwo jest minimalistyczne” - mniej znaczy lepiej – ponieważ dobre wzornictwo koncentruje się na najważniejszych aspektach, produkty nie są błahe. Powrót do czystości, powrót do prostoty [gizmodo.com, 2010].

W kontekście systemów BI, są to bardzo ogólne reguły, które nie zawsze przekładają się na proste porady w jaki sposób projektować raporty. Mając jednak na uwadze uniwersalność wymienionych cech oraz wyjątkowo szybki postęp technologiczny obecnych czasów, warto korzystać z nich w formie inspiracji dla każdego projektu.

Niezmiennie ważnym aspektem w tworzeniu raportów jest tzw. percepcja wizualna człowieka. Różnica między „przedświadomym” a świadomym przetwarzaniem informacji jest łatwa do zaprezentowania. W ramach doświadczenia, możemy posłużyć się przykładem na znalezienie wszystkich cyfr 5 w przykładach 1 oraz 2.

Przykład 1:

51324130295209480920409009899

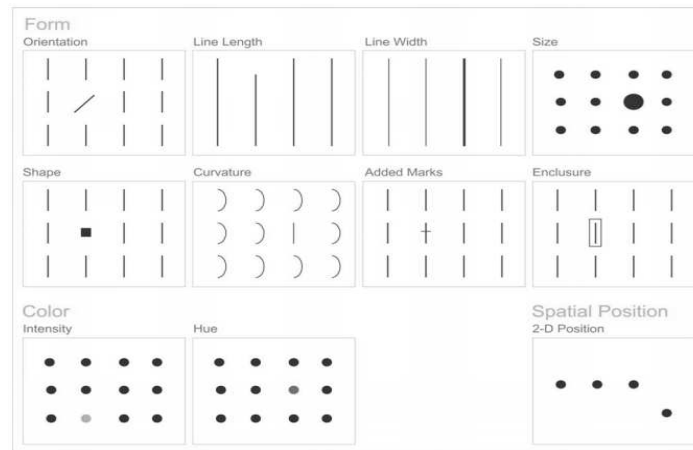
Przykład 2:

5132413029**5**209480920409009899

Prawidłowa odpowiedź brzmi: w przykładach występują dwie piątki. Szybkie odnalezienie szukanej liczby było możliwe dzięki odznaczającym się piątkom w przykładzie 2, co nie umknęło „przedświadomości” percepcji wizualnej. Warto zauważyć, iż w przykładzie pierwszym, „przedświadomość” po nieudanym odnalezieniu cech szczególnych, została kompletnie pominięta i mózg przeszedł do świadomego etapu analizy w poszukiwaniu piątek cyfra po cyfrze. Oprócz koloru, istnieją inne atrybuty ułatwiające „przedświadomości” odszukiwanie podobizn między obiektami. Colin Ware opisuje następujące atrybuty oddziałujące na „przedświadomość” percepcji wizualnej:

- kolor, na który składa się jego intensywność oraz odcień,
- kształt, na który składa się jego orientacja, długość, szerokość, wielkość, kształt, krzywizna, oznaczenia oraz otoczenie,
- pozycja w przestrzeni,
- ruch [Ware 2000].

Rysunek 2. Atrybuty „przedświadomej” percepcji wizualnej



Źródło: [Few 2004]

Każdy z wymienionych atrybutów może być świadomie wykorzystany do pogrupowania bądź wyróżnienia informacji w raporcie.

## PODSUMOWANIE

Wizualizacja danych, a raczej przedstawienie z terabajtów, czasem już petabajtów, danych istotnych informacji to klucz do sukcesu każdej firmy. Jeszcze 20 lat temu nie miało to aż tak wielkiego znaczenia jak obecnie. Dlatego też autorzy niniejszego artykułu postanowili zaprezentować najczęstsze trendy w tworzeniu raportów: dashboardów i balanced scorecardów. W ogólnym rozrachunku, niezależnie od tego, do czego dany raport czy też wykres służy, najważniejszą zasadą jego prezentacji jest dążenie do jak najbardziej prostej, a przy tym przejrzystej formy. Parafrazując Ditera Ramsa, możemy to zamknąć w czterech słowach „dobre wzornictwo jest minimalistyczne”.

## BIBLIOGRAFIA

- Diaz J. (2008) 1960s Braun Products Hold the Secrets to Apple's Future. gizmodo.com.  
 Few S. (2004) Tapping the Power of Visual Perception. Perceptual Edge, 3-5.  
 Few S. (2013) Information Dashboard Design. O'Reilly.  
 Few S. (2014) Display Platforms for Quantitative Information. Perceptual Edge.  
 Jelen B. (2005) The Spreadsheet at 25: 25 The Evolved of the Invention that Changed the World. Holy Macro! Books.  
 Kaplan R. S., Norton D. P. (1991) The Balanced Scorecard: Measures That Drive Performance. Harvard Business Review, 71-79.  
 Michael A., Jelen B. (2005) Pivot Table Data Crunching (QUE).



- Schneiderman A. M. (1990) Pilot Demo- presentation.
- Shivakumar B. L. (2014) A Review of Trends and Technologies in Business Analytics International Journal of Advanced Research in Computer Science, 5 (8).
- United Nations Economic Commission for Europe Making Data Meaningful Part 2: A guide to presenting statistics (2009).
- Ware C. (2000) Information Visualization: Perception for Design. Academic Press.
- Watson H., Walls J. (1993) Executive Information systems. System Sciences, Proceeding of the Twenty-Sixth Hawaii International Conference on System Sciences.
- [www.bi.pl](http://www.bi.pl).
- [www.codecdss.co.uk](http://www.codecdss.co.uk).
- [www.en.wikipedia.org/wiki/Lotus\\_Improv](http://www.en.wikipedia.org/wiki/Lotus_Improv).
- [www.en.wikipedia.org/wiki/Dieter\\_Rams](http://www.en.wikipedia.org/wiki/Dieter_Rams).
- [www.en.oxforddictionaries.com](http://www.en.oxforddictionaries.com).
- [www.pl.wikipedia.org/wiki/Enron](http://www.pl.wikipedia.org/wiki/Enron).

## PRESENTATION METHODS OF DATA IN BI REPORTS

**Abstract:** In the XXI century we are dealing with large amount of data which we have to process. The pressure to achieve the biggest amount of information is on the top level. Very huge problem is filtration of the most significant information from this data stream. To solve above problem author of this article undertakes request to present main principles of data presentation in new Business Intelligence systems. The main role of assumed principles is readability and interpretation of presented data improvement.

**Keywords:** BI reports, data presentation, dashboards visualization

## **POLISH REVIEWERS COOPERATING WITH THE JOURNAL IN 2017**

Elżbieta Badach	Karol Kukula
Iwona Bąk	Tadeusz Kwater
Beata Bieszk-Stolorz	Jacek Kwiatkowski
Agata Boratyńska	Joanna Landmesser
Bolesław Borkowski	Sławomir Lisek
Jadwiga Bożek	Lidia Luty
Mariola Chrzanowska	Piotr Łukasiewicz
Hanna Dudek	Arkadiusz Manikowski
Krzysztof Gajowniczek	Jerzy Marzec
Stanisław Gędek	Aleksandra Matuszewska-Janica
Michał Gostkowski	Mariola Piłatowska
Wiesław Grygierzec	Artur Prędko
Monika Jaworska	Jacek Strojny
Stanisław Jaworski	Ewa M. Syczewska
Alina Jędrzejczak	Wiesław Szczesny
Andrzej Karpio	Janina Szewczyk
Marek Karwański	Mirosław Wasilewski
Joanna Kisielińska	Ewa Wędrowska
Marek Kociński	Dorota Witkowska
Krzysztof Kompa	Janusz Wywiół
Grzegorz Koszela	Tomasz Ząbkowski
Marek Kośny	Rafał Zbyrowski
Barbara Kowalczyk	Monika Zielińska-Sitkiewicz
Monika Krawiec	Wojciech Zieliński
Justyna Kujawska	Monika Ziolo

## **FOREIGN REVIEWERS COOPERATING WITH THE JOURNAL IN 2017**

Paolo Gajo (University of Florence, Italy)  
Vasile Glavan (Moldova State University, Moldova)  
Yuriy Kondratenko (Black Sea National University, Ukraine)  
Vassilis Kostoglou (Alexander Technological Educational Institute of Thessaloniki, Greece)  
Nicholas N. Olenev (Russian Academy of Sciences, Russia)  
Alexander N. Prokopenya (Brest State Technical University, Belarus)  
Yochanan Shachmurove (The City College of The City University of New York, USA)  
Mirbulat B. Sikhov (al-Farabi Kazakh National University, Kazakhstan)