Cena 12,00 zł
(VAT 5%)

Indeks 381306
PL ISSN 0043-518X
e-ISSN 2543-8476

WIADOMOŚCI STATYSTYCZNE

THE POLISH STATISTICIAN

GŁÓWNY URZĄD STATYSTYCZNY
STATISTICS POLAND

POLSKIE TOWARZYSTWO STATYSTYCZNE
POLISH STATISTICAL ASSOCIATION

MIESIĘCZNIK
MONTHLY JOURNAL
WRZESIEŃ 2019
SEPTEMBER

Numer 9 (700)
Issue

GUS

TPS

Cena 12,00 zł
(VAT 5%)

Indeks 381306
PL ISSN 0043-518X
e-ISSN 2543-8476

WIADOMOŚCI STATYSTYCZNE

THE POLISH STATISTICIAN

GŁÓWNY URZĄD STATYSTYCZNY
STATISTICS POLAND

POLSKIE TOWARZYSTWO STATYSTYCZNE
POLISH STATISTICAL ASSOCIATION

MIESIĘCZNIK
MONTHLY JOURNAL
ROK LXIV
VOLUME 64
WRZESIEŃ
SEPTEMBER 2019

Numer 9 (700)
Issue

ADRES REDAKCJI / EDITORIAL OFFICE ADDRESS

GUS, al. Niepodległości 208, 00-925 Warszawa, tel./phone +48 22 608 32 25
e-mail: redakcja.ws@stat.gov.pl

Wersja elektroniczna, stanowiąca wersję pierwotną czasopisma, jest dostępna na ws.stat.gov.pl
An electronic edition of the journal is an original one. It is available at ws.stat.gov.pl

Zakład Wydawnictw Statystycznych

Zakład Wydawnictw Statystycznych / Statistical Publishing Establishment
al. Niepodległości 208, 00-925 Warszawa, tel./phone +48 22 608 31 45
Informacje w sprawie nabywania czasopism / Information on purchasing of the journal
tel./phone +48 22 608 32 10, +48 22 608 38 10
Zbigniew Karpiński (redaktor techniczny / technical editor)
Ewa Krawczyńska (skład i łamanie / typesetting)
Wydział Korekty pod kierunkiem Bożeny Gorczycy / Proof-Reading Section supervised by Bożena Gorczyca
Andrzej Kajkowski (wykresy/figures)

Indeks 381306

Prenumerata jest prowadzona przez / Subscription is realised by RUCH S.A.
Zamówienia na prenumeratę można składać na stronie / Orders at www.prenumerata.ruch.com.pl

Zam. 320/2019 – nakład 340 egz.

Spis treści
Contents

Od redakcji

W numerze „Wiadomości Statystycznych. The Polish Statistician", który oddajemy w ręce czytelników, zamieszczamy prace dotyczące zarówno rozwiązań metodologicznych, jak i zastosowań narzędzi statystycznych w praktyce, a także artykuły podejmujące wyzwania badawcze.

Wydanie otwiera artykuł Grzegorza Przekoty *Wymiar fraktalny szeregów czasowych szacowanych metodą podziału pola*. Badacz rozwija obiecującą autorską metodę szacowania wymiaru fraktalnego – porusza kwestię określenia zmienności oraz identyfikacji procesu kształtowania się wartości szeregów czasowych, a w głównej mierze przewidywalności tych zmian przy zastosowaniu wymiaru fraktalnego. Za jego pomocą autor opisuje właściwości szeregu czasowego, wartości indeksu giełdowego WIG w latach 2014–2018 oraz szeregów czasowych stóp wzrostu największych polskich spółek giełdowych w latach 2015–2018.

W artykule *Zmiana sektorowej struktury pracujących w przekroju województw* Katarzyna Wawrzyniak i Barbara Batóg podejmują próbę zidentyfikowania prawidłowości w zakresie zmian w strukturze pracujących według sektorów ekonomicznych – ważnego zagadnienia, ściśle związanego z rozwojem gospodarczym. Autorki porównują stopień podobieństwa struktury pracujących w poszczególnych województwach na tle kraju. Oszacowują również zróżnicowanie udziału procentowego pracujących według sektorów, wykorzystując wskaźnik Herfindahla-Hirschmana, oraz badają podobieństwo struktury pracujących w ujęciu łańcuchowym.

Dariusz Kotlewski w artykule *Dekompozycje czynnikowe przyrostu wartości dodanej brutto według sekcji PKD i województw* przedstawia zalety dekompozycji przyrostu wartości dodanej brutto w kilku wariantach. Takie podejście pozwala na bardziej szczegółową analizę procesów gospodarczych zachodzących w kraju. Na uwagę zasługuje użyteczność metodologii opracowanej przez autora, która umożliwia przeprowadzenie obliczeń jednocześnie na poziomie sekcji rodzaju działalności (według PKD) i województw, co jest bardzo przydatne w analizie danych w ujęciu regionalnym.

Korzyści ze stosowania narzędzi do automatycznego zbierania danych oraz związane z tym wyzwania to przedmiot artykułu *Pozyskiwanie i analiza danych na temat ofert pracy z wykorzystaniem big data* Jacka Maślankowskiego. Autor prezentuje wyniki eksperymentalnych badań opartych na analizie danych pochodzących z najpopularniejszych portali z ofertami pracy, pozyskanych za pomocą metod web scrapingu oraz text miningu.

W pierwszą rocznicę śmierci wspominamy dr. Dariusza Parysa. O pracy dydaktycznej i dorobku naukowym tego uznanego statystyka pisze Czesław Domański.

Na zakończenie nawiązujemy do dyskusji nad możliwościami wykorzystania w statystyce nowych danych, jaka toczyła się na konferencji Metodologia Badań Statystycznych MET2019. W sprawozdaniu pokonferencyjnym Daniel Koprowicz omawia najciekawsze referaty; uwagę poświęca także warsztatom prowadzonym przez prof. Grahama Kaltona, które dotyczyły metod próbkowania populacji.

Zapraszamy do lektury.

From the editorial team

The current issue of *Wiadomości Statystyczne. The Polish Statistician* features articles focused on methodological solutions, applications of statistical tools in practice as well as papers undertaking research challenges.

Fractal dimension of time series estimated by the surface division method, an opening paper by Grzegorz Przekota, presents a promising original method for estimating the fractal dimension. The author addresses issues such as determining the variability of time series and identifying the process of shaping their values, and, to the largest extent, determining the predictability of these changes with the use of the fractal dimension. This method
is used by the author to describe the properties of the time series of the values of WIG stock exchange index in 2014–2018 and the properties of the time series of the growth rates of the Polish largest listed companies in 2015–2018.

In Katarzyna Wawrzyniak and Barbara Batóg's article entitled: *The change in the sectoral structure of employed persons in voivodships*, the authors attempt to identify the regularities in the changes of the employment structure by economic sectors, an important issue directly related to economic development. They compare the degree of similarity between the sectoral structure of employment of each examined voivodship and such structure of whole Poland. They also estimate the diversification of the percentages of persons employed in particular sectors within each voivodship using the Herfindahl-Hirschman relative index, as well as examining the similarities between sectoral structures of employment using the chain approach.

Factor decompositions of gross value added growth by the NACE sections and voivodships by Dariusz Kotlewski presents the merits of gross value added growth decompositions in several variants. The author's approach makes it possible to analyse economic processes in the Polish economy in detail greater than before. Thanks to his methodology, it is possible to perform calculations according to the NACE sections and according to voivodships at the same time, which proves very useful, especially in analysing regional data.

The benefits of using web scraping methods to gather data and the challenges connected to this process are presented in Jacek Maślankowski's article entitled: *The collection and analysis of the data on job advertisements with the use of big data*. His paper presents the results of experimental research based on the analysis of data from the most popular job-searching websites, obtained by web-scraping and text-mining methods.

We are also reminiscing about the late Dr Dariusz Parys on the first anniversary of his death. Didactic and scientific achievement of this recognised Polish statistician is presented by Czesław Domański.

The issue concludes with the reference to a discussion, held at the Methodology of Statistical Research MET2019 conference, about the possibilities of using new data in statistics. Daniel Koprowicz presents the most interesting speeches and presentations from the conference, as well as the workshops exploring the methods for sampling population led by Prof Graham Kalton.

We wish you a pleasant reading.

Wiadomości Statystyczne. The Polish Statistician, 2019, vol. 64, 9, 7–24 DOI: 10.5604/01.3001.0013.7594
Studia metodologiczne / Methodological studies

Wymiar fraktalny szeregów czasowych szacowany metodą podziału pola

Grzegorz Przekota[a]

Streszczenie. Jedną z ważniejszych kwestii do rozstrzygnięcia w analizie szeregów czasowych jest określenie ich zmienności oraz identyfikacja procesu kształtowania ich wartości. W ujęciu klasycznym zmienność najczęściej utożsamiana jest z wariancją stóp wzrostu. Tymczasem natura ryzyka to nie tylko zmienność, lecz także przewidywalność zmian, którą można ocenić przy użyciu wymiaru fraktalnego. Celem artykułu jest prezentacja zastosowania wymiaru fraktalnego szacowanego metodą podziału pola do oceny właściwości szeregów czasowych. W opracowaniu przedstawiono sposób wyznaczenia wymiaru fraktalnego, jego interpretację, tablice istotności oraz przykład zastosowania. Za pomocą wymiaru fraktalnego opisano właściwości szeregu czasowego wartości indeksu giełdowego WIG w latach 2014–2018 oraz szeregów czasowych stóp wzrostu największych polskich spółek giełdowych w latach 2015–2018. Zastosowana metoda umożliwia zakwalifikowanie szeregu czasowego do jednej z trzech klas, jako szereg: persystentny, błądzenia losowego bądź antypersystentny. Na szczególnych przypadkach pokazano różnice pomiędzy zastosowaniem odchylenia standardowego i wymiaru fraktalnego do oceny ryzyka. Wymiar fraktalny jawi się tu jako metoda pozwalająca na ocenę stopnia stabilności wahań.

Słowa kluczowe: wymiar fraktalny, metoda podziału pola, zmienność, błądzenie losowe

Fractal dimension of time series estimated by the surface division method

Summary. One of the most important issues to be settled in the analysis of time series is determining their variability and identifying the process of shaping their values. In the classical approach, volatility is most often identified with the variance of growth rates. However, risk can be characterised not only by the variability, but also by the predictability of the changes which can be evaluated using the fractal dimension. The aim of this paper is to present the applicability of the fractal dimension estimated by the surface division method to the assessment of the properties of time series. The paper presents a method for determining the fractal dimension, its interpretation, significance tables and an example of its application. Fractal dimension has been used here to describe the properties of the time series of the WIG stock exchange index in 2014–2018 and the time series of the growth rates of the largest listed Polish companies in 2015–2018. The applied method makes it possible to classify a time series into one of three classes of series: persistent, random or antipersistent. Specific cases show the differences between the use of standard deviation and fractal dimension for risk assessment. Fractal dimension appears here to be a method for assessing the degree of stability of variations.

Keywords: fractal dimension, surface division method, variability, random walk

JEL: C18, C22

[a] Politechnika Koszalińska, Wydział Nauk Ekonomicznych.

Współczesne procesy gospodarcze nie dają się wyjaśnić za pomocą klasycznych metod statystyki czy podstawowych praw ekonomii. Na zjawiska takie, jak: występowanie cykli koniunkturalnych, sytuacja na rynku pracy, kształtowanie stopy procentowej czy cen na rynkach finansowych wpływa wiele zmiennych. Ich liczba oraz skomplikowane powiązania między nimi sprawiają, że proste modele analityczne okazują się w praktyce nieskuteczne. Większe możliwości daje modelowanie wielorównaniowe ze zmiennymi opóźnionymi, choć i ono bywa przedmiotem krytyki. Peters (1997) wskazuje, że prognozy z tradycyjnych modeli ekonometrycznych nie sprawdzają się w praktyce lub też sprawdzają się tylko w perspektywie krótkiego czasu. Ponadto niewielka zmiana warunków początkowych powoduje, że model przestaje funkcjonować poprawnie. Problemem jest także założenie równowagi systemu gospodarczego, nieprzystające do rzeczywistości go otaczającej, która ewoluuje, powodując także zmiany samego systemu. Charakterystyczny dla systemu gospodarczego będzie zatem raczej stan nierównowagi niż równowagi.

Rozwój nowych dyscyplin naukowych, takich jak teoria chaosu i analiza fraktalna, daje szersze możliwości opisywania otaczającej rzeczywistości. Do modelowania zjawisk ekonomicznych można wykorzystać szereg stosunkowo prostych systemów chaosu deterministycznego: równanie logistyczne, atraktor Henona, będący dwuwymiarowym odpowiednikiem równania logistycznego (Mosdorf, 1997), czy też model Lorenza (Zawadzki, 1996), znany jako efekt motyla. Popularnym narzędziem wywodzącym się z geometrii fraktalnej jest także wymiar fraktalny.

Jeżeli system jest nieliniowym systemem dynamicznym, to charakteryzują go występowanie długoterminowych korelacji i trendów, nieoczekiwane zachowanie w pewnych warunkach i pewnych okresach oraz struktura, której części (zarówno odcinki większe, jak i coraz mniejsze) są podobne do całości i mają te same charakterystyki statystyczne. Taką strukturę określa się jako fraktalną (Siemieniuk, 2001).

Celem artykułu jest prezentacja zastosowania wymiaru fraktalnego szacowanego metodą podziału pola do oceny zmienności szeregów czasowych. W opracowaniu zwrócono uwagę na interpretację uzyskiwanych wartości oraz przedstawiono tabelę istotności. Pokazano różnice w postrzeganiu ryzyka rozumianego jako zmienność i mierzonego odchyleniem standardowym oraz ryzyka rozumianego jako stabilność wahań i mierzonego wymiarem fraktalnym. Praktyczne zastosowanie i ograniczenia metody omówiono na przykładzie szeregów czasowych poziomów i zmian wartości indeksu giełdowego WIG oraz notowań największych polskich spółek giełdowych. Zakres czasowy badania to cztery lata, od października 2014 r. do października 2018 r. Alternatywnie dokonano porównania uzyskanych wyników z wynikami testu stacjonarności ADF.

WYMIAR FRAKTALNY SZEREGÓW CZASOWYCH

W klasycznej teorii inwestycji finansowych jedną z najpopularniejszych miar określających ryzyko jest wariancja stóp zwrotu. Uważa się, że ryzyko jest tym większe, im większa zmienność stóp zwrotu. Jednak w badaniu ryzyka systemów o charakterze losowym odpowiednią miarę stanowi wariancja. Interesujących rozwiązań w zakresie szacowania ryzyka dostarcza geometria fraktalna. Jedną z miar wywodzących się z geometrii fraktalnej jest wymiar fraktalny szeregów czasowych, który może być uzupełnieniem klasycznych miar zmienności. Chociaż w literaturze przedmiotu bardzo często wymiar fraktalny traktuje się jako miarę ryzyka w rozumieniu zmienności (Zeug-Żebro, 2015), to jego natura jest inna. Problem ten poruszono w niniejszej pracy, porównując wartości wymiaru fraktalnego oraz odchylenia standardowego dla pewnych szczególnych przypadków.

Mandelbrot (1982) podaje przykład zastosowania wymiaru fraktalnego do analizy zjawisk naturalnych. Problem dotyczy pomiaru długości linii brzegowej. Wynik zależy od długości miarki użytej do pomiaru: im miarka jest krótsza, tym wynik dokładniejszy, gdyż pozwala na uchwycenie większej liczby krzywizn. Wymiar fraktalny daje odpowiedź na pytanie, jak postrzępione są linie brzegowe (im bardziej, tym wymiar ten jest większy). Przykładowo wymiar fraktalny linii brzegowej Norwegii wynosi 1,52, a linii brzegowej Wielkiej Brytanii – 1,26. Wynik ten jest zgodny ze spostrzeżeniami poczynionymi na podstawie mapy, ponieważ linia brzegowa Norwegii jest bardziej postrzępiona od linii brzegowej Wielkiej Brytanii, a więc jej wymiar fraktalny jest większy i bardziej zbliżony do 2.

Współcześnie wymiar fraktalny stosuje się do opisu wielu zjawisk naturalnych (Cervantes-De la Torre, González-Trejo, Real-Ramírez i Hoyos-Reyes, 2013), zagospodarowania przestrzennego (Chen, 2013; Guoqiang, 2002), problemów medycznych (Gómez, Mediavilla, Hornero, Abásolo i Fernández, 2009; Harne, 2014) czy ekonomicznych (Bhatt, Dedania i Shah, 2015; Kapecka, 2013). Rozwijana jest także metodologia szacowania wymiaru fraktalnego (Sy-Sang i Feng-Yuan, 2009).

W polskiej literaturze przedmiotu ostatnich lat wiele uwagi wymiarowi fraktalnemu poświęcił Buła (2017). Porównywał on m.in. wyniki otrzymane za pomocą prezentowanej w tym opracowaniu metody podziału pola oraz za pomocą innych metod. Okazało się, że podział pola daje niższe wartości niż inne metody. Nie stanowi to jednak problemu, ponieważ kluczową kwestią jest nie wartość, lecz rozstrzygnięcie, czy dany proces jest procesem błądzenia losowego, czy też innym.

Niniejsza praca porusza problem testowania hipotezy błądzenia losowego przy użyciu metody podziału pola. Szacowanie wymiaru fraktalnego dla ekonomicznych szeregów czasowych wymaga odejścia od klasycznej geometrii euklidesowej, podającej wymiar przestrzeni, w której umieszczony jest wykres szeregu czasowego. Przestrzeń tę stanowi płaszczyzna o wymiarze euklidesowym 2.

Rozpatrując natomiast trajektorię szeregu czasowego jako łamaną, otrzymujemy wymiar euklidesowy 1. Tymczasem wykres szeregu czasowego nie wypełnia całej płaszczyzny, na której został umieszczony, zatem jego wymiar będzie mniejszy od 2 (wymiaru euklidesowego płaszczyzny) i większy od 1, ponieważ jest to wymiar euklidesowy linii prostej, a szeregi czasowe na ogół mają inny kształt.

Wymiar fraktalny jako wymiar ułamkowy pozwala charakteryzować kształt wykresu szeregu czasowego. Opisuje, w jaki sposób szereg czasowy wypełnia swoją przestrzeń i jest wynikiem wszystkich czynników wpływających na system, z którego pochodzi szereg czasowy (Peters, 1997). Efektem oddziaływania różnych czynników może być obraz szeregów czasowych zmiennych ekonomicznych, który będzie klasyfikował je do jednej z trzech klas różniących się wartościami wymiaru fraktalnego (Halley i Kunin, 1999):

1. Szeregi persystentne (czarny szum) – w których obecne jest zjawisko wzmacniania trendu. Oznacza to (w przeciwieństwie do szeregów antypersystentnych), że jeżeli wartość szeregu wzrosła (lub spadła) w porównaniu z wartością poprzednią, to w następnym momencie bardziej prawdopodobny będzie jej wzrost (lub spadek). Szeregi takie będą miały wymiar fraktalny bliższy 1;
2. Szeregi czasowe błądzenia losowego (biały szum) – w których poprzednia zmiana wartości nie ma wpływu na przyszłe zmiany. Zdarzenia wpływające na wartości szeregu są przypadkowe i nieskorelowane ze sobą, stąd też szeregi takie są nieprzewidywalne;
3. Szeregi antypersystentne (różowy szum) – w których występuje zjawisko powracania wartości obserwacji do poziomu średniego. Oznacza to, że jeśli wartość szeregu odchyli się w górę (lub w dół) od średniej, to w następnym momencie bardziej prawdopodobne jest odchylenie w przeciwnym kierunku. Szeregi takie będą miały wymiar fraktalny bliższy 2.

Wymiar fraktalny dla wykresów jednowymiarowych szeregów czasowych przyjmuje wartości z przedziału [1, 2]: 1 – gdy wykres przybierze kształt linii prostej, 2 – gdy wypełni pewien obszar dwuwymiarowy na płaszczyźnie. W praktyce wartości skrajne nie są osiągane.

Skoro wymiar fraktalny ma opisywać, jak szereg czasowy wypełnia obszar, lub inaczej – jak zagęszcza się na płaszczyźnie, to większe zagęszczenie będzie powodować zwiększenie wymiaru fraktalnego. Oznacza to, że częste zmiany szeregów czasowych w różnych kierunkach powodują zwiększenie wymiaru fraktalnego i większe wypełnienie płaszczyzny. Występuje tu zjawisko powrotu do średniej. Szeregi jednokierunkowe, z małą liczbą zmian, mają mniejszy wymiar fraktalny, ich kształty zaś są bardziej zbliżone do kształtu prostej. Takie szeregi charakteryzuje zjawisko podtrzymania trendu.

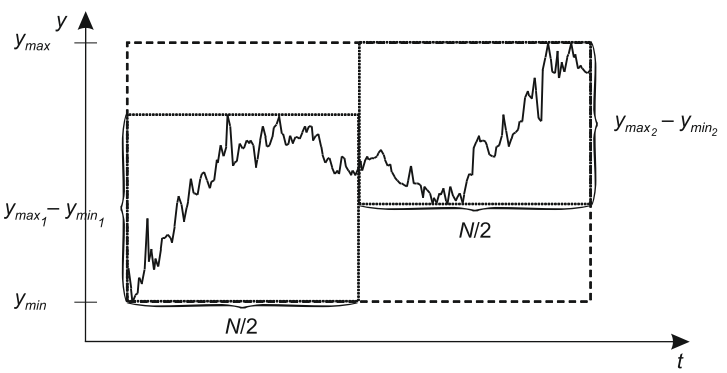
Do oszacowania wymiaru fraktalnego ekonomicznych szeregów czasowych można wykorzystać metodę pudełkową BCM, w której zamiast kół zlicza się kwadraty o boku określonej długości, potrzebne do pokrycia wykresu szeregu

czasowego. Inną popularną metodą szacowania wymiaru fraktalnego jest metoda wariacyjna VM (Dubuc, Quiniou, Roques-Carmes, Tricot i Zucker, 1989), podobnie jak metoda segmentowo-wariacyjna SVM (Zwolankowska, 2001). Często stosuje się także analizę przeskalowanego zakresu autorstwa Hursta (Hurst, 1951; Kale i Butar Butar, 2011).

METODA PODZIAŁU POLA

Metoda szacowania wymiaru fraktalnego przedstawiona w opracowaniu łączy w sobie elementy metody segmentowo-wariacyjnej oraz tradycyjnych metod geometrycznych (Przekota i Przekota, 2004). Podobnie jak w metodzie segmentowo-wariacyjnej wykres szeregu czasowego pokrywany jest przez prostokąty. Samo szacowanie wymiaru fraktalnego odbywa się natomiast poprzez szacowanie współczynnika regresji, tak jak w metodach geometrycznych. W kolejnych punktach wyprowadzono wzór wymiaru *MPP* (metoda podziału pola) oraz pokazano jego zastosowania praktyczne.

WYKR. 1. SZEREG CZASOWY NA PŁASZCZYŹNIE



Ź r ó d ł o: opracowanie własne.

Niech szereg czasowy (wykr. 1) ma długość *N*. Wtedy pole obszaru zajmowanego przez szereg można definiować jako:

$$P = N \cdot (y_{max} - y_{min}) \tag{1}$$

gdzie:
y_{max} – największa wartość w szeregu,
y_{min} – najmniejsza wartość w szeregu.

Po podziale szeregu na połowy pole wyrażać będzie się wzorem:

$$p = \frac{N}{2} \cdot \left(y_{max_1} - y_{min_1}\right) + \frac{N}{2} \cdot \left(y_{max_2} - y_{min_2}\right) \tag{2}$$

Pomiędzy *p* a *P* zachodzi nierówność:

$$p \leq P \tag{3}$$

Przy powtarzaniu czynności przepoławiania kolejnych fragmentów szeregu za każdym razem okazuje się, że suma pól po podziale nie jest większa od sumy pól pierwotnych. Oznacza to, że przy dowolnym podziale pierwotnym na *k* części pole zajmowane przez wykres szeregu będzie wynosić:

$$P_k = \frac{N}{k}\sum_{i=1}^{k}\left(y_{max_i} - y_{min_i}\right) \tag{4}$$

a przy podziale na 2*k* części:

$$P_{2k} = \frac{N}{2k}\sum_{i=1}^{2k}\left(y_{max_i} - y_{min_i}\right) \tag{5}$$

Pomiędzy P_k i P_{2k} zachodzi nierówność:

$$P_{2k} \leq P_k \tag{6}$$

co można także zapisać jako:

$$P_{2k} \leq 2 \cdot \frac{P_k}{2} \tag{7}$$

Dla dowolnego szeregu natomiast zachodzi:

$$P_{2k} = MPP_k \cdot \frac{P_k}{2} \tag{8}$$

MPP_k zawiera się w przedziale [1, 2] i będzie tym większe, im kształt trajektorii szeregu czasowego będzie bardziej postrzępiony, czyli im częściej w szeregu wystąpi zmiana trendu na przeciwny. Natomiast im bardziej kształt szeregu zbliży się do prostej, czyli im mniej zmian trendu na przeciwny wystąpi w szeregu, tym bardziej wartość MPP_k będzie bliższa 1.

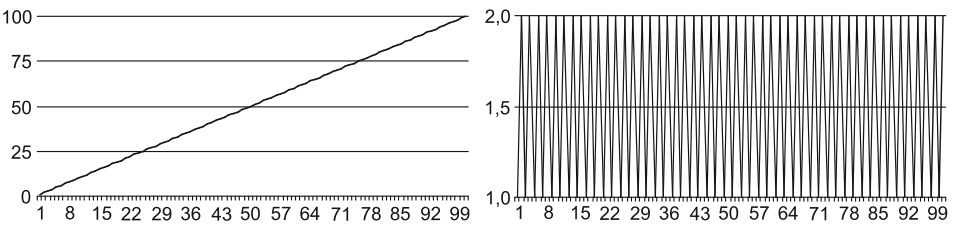
Jeśli w układzie współrzędnych na osi *x* będą odkładać się wartości $\frac{P_k}{2}$, a na osi *y* wartości P_{2k}, to wartość MPP_k będzie współczynnikiem regresji liniowej *y* względem *x* bez wyrazu wolnego, gdzie wartości P_{2k} odgrywają rolę zmiennej objaśnianej, a wartości $\frac{P_k}{2}$ – zmiennej objaśniającej. Stąd:

$$MPP_k = \frac{\sum_k P_{2k}\frac{P_k}{2}}{\sum_k \left(\frac{P_k}{2}\right)^2} \tag{9}$$

gdzie *k* – liczba dokonanych podziałów.

Tak zdefiniowana wartość MPP_k może być traktowana jako miara postrzępienia szeregów, czyli jako wymiar fraktalny szeregów. W praktycznym zastosowaniu dla danych z rynków finansowych na podstawie wartości wymiaru fraktalnego można wnioskować o ryzyku inwestycyjnym.

WYKR. 2. SKRAJNE PRZYPADKI WYMIARU FRAKTALNEGO *MPP*



Ź r ó d ł o: opracowanie własne.

Dwa skrajne przypadki wymiaru fraktalnego *MPP* pokazano na wykr. 2. Pierwszy to linia prosta (tutaj jest to funkcja $y = x$). Pole po podziale zawsze stanowi połowę pola pierwotnego, a zatem:

$$P_{2k} = 1 \cdot \frac{P_k}{2} \tag{10}$$

czyli wymiar fraktalny jest tutaj równy 1 i jest zgodny z wymiarem euklidesowym prostej.

Drugi przypadek to sytuacja, w której wartości szeregu na przemian rosną i maleją (tutaj jest to 2 dla x parzystych i 1 dla x nieparzystych). Pole po podziale zawsze jest równe polu pierwotnemu, zatem:

$$P_{2k} = 2 \cdot \frac{P_k}{2} \tag{11}$$

czyli wymiar fraktalny jest tutaj równy 2 i jest zgodny z wymiarem euklidesowym płaszczyzny.

W celu rozróżnienia szeregów błądzenia losowego i szeregów persystentnych i antypersystentnych przeprowadzono symulacje Monte Carlo wymiaru fraktalnego szeregów błądzenia losowego. W tabl. 1 zawarto statystyki tych symulacji wraz z wynikami testów normalności rozkładu uzyskanych wartości. Łącznie przeprowadzono 800 symulacji, po 100 w każdej próbie.

TABL. 1. STATYSTYKI WYMIARU FRAKTALNEGO SYMULOWANYCH SZEREGÓW BŁĄDZENIA LOSOWEGO METODĄ MONTE CARLO

Wyszczególnienie	*N* = 200		*N* = 500		*N* = 1000		*N* = 1600	
	próby							
	1	2	1	2	1	2	1	2
$\bar{x}$	1,3707	1,3731	1,3788	1,3809	1,3843	1,3790	1,3765	1,3703
S	0,1183	0,1140	0,1127	0,1145	0,1100	0,1182	0,1112	0,1017
Minimum	1,1479	1,1485	1,1342	1,1549	1,1580	1,1347	1,1492	1,1573
Maksimum	1,6568	1,6594	1,6160	1,6372	1,6125	1,6151	1,6106	1,6041
TN Kołmogorowa--Smirnowa	$p > 0{,}20$							
TN Lillieforsa	$p > 0{,}20$							
TN Shapiro-Wilka ..	$p = 0{,}1421$	$p = 0{,}1591$	$p = 0{,}2305$	$p = 0{,}3516$	$p = 0{,}1939$	$p = 0{,}1716$	$p = 0{,}2338$	$p = 0{,}4346$

U w a g a. $\bar{x}$ – średnia arytmetyczna, *S* – odchylenie standardowe, TN – test normalności, *p* – poziom istotności.
Ź r ó d ł o: obliczenia własne.

Na podstawie uzyskanych wyników testów normalności (tabl. 1) stwierdzono, że rozkład wymiaru fraktalnego jest normalny, o średniej i odchyleniu standardowym przyjętych z uśrednionych procesów symulowanych. Na tej podstawie skonstruowano tablice istotności wymiaru *MPP*. Dane te dla wybranych poziomów istotności zaprezentowano w tabl. 2. Weryfikacji podlega hipoteza zerowa: proces generujący szereg czasowy jest procesem błądzenia losowego.

TABL. 2. GRANICE ODRZUCEŃ HIPOTEZY BŁĄDZENIA LOSOWEGO SZEREGÓW CZASOWYCH DLA WYMIARU *MPP*

N		Poziom istotności α		
a – dolna granica b – górna granica		0,2	0,1	0,05
200	a	1,2234	1,1813	1,1447
	b	1,5204	1,5626	1,5991
500	a	1,2347	1,1935	1,1578
	b	1,5250	1,5662	1,6019
1000	a	1,2357	1,1942	1,1584
	b	1,5277	1,5691	1,6050
1600	a	1,2371	1,1985	1,1650
	b	1,5097	1,5484	1,5819

Ź r ó d ł o: obliczenia własne.

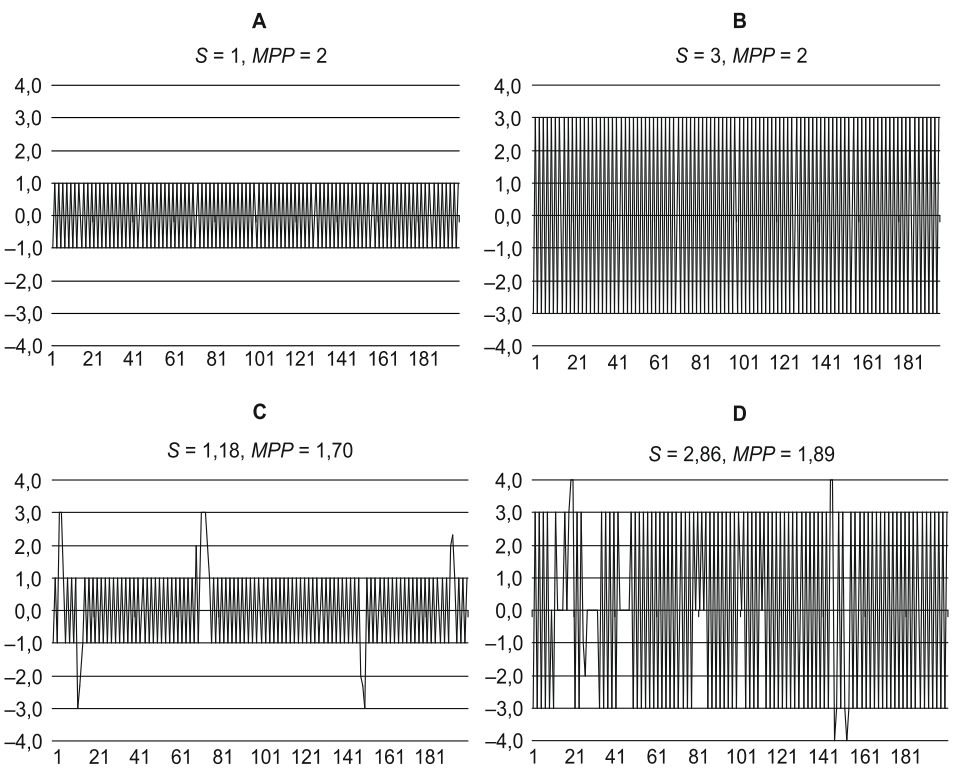
W zależności od wartości *MPP* można wyróżnić trzy klasy szeregów czasowych:

1. Wartości *MPP* poniżej dolnej granicy oznaczają szeregi persystentne, a więc procesy ze wzmacnianiem trendu. Są to szeregi przewidywalne;

2. Wartości *MPP* pomiędzy dolną a górną granicą oznaczają szeregi, w których przebieg może być kształtowany przez procesy błądzenia losowego. Są to szeregi nieprzewidywalne;
3. Wartości *MPP* powyżej górnej granicy oznaczają szeregi antypersystentne, a więc procesy charakteryzowane przez zjawisko powrotu do wartości średniej. Są one (tak jak szeregi persystentne) przewidywalne.

Wymiar *MPP*, w porównaniu z odchyleniem standardowym, ma interesujące właściwości w ocenie zmienności szeregów stóp wzrostu. Natura tych szeregów jest antypersystentna, jednak zdarzają się pewne nietypowe sytuacje. Warto zobaczyć, jak w takiej sytuacji zachowuje się odchylenie standardowe i wymiar *MPP*. Pewne szczególne przypadki ujęto na wykr. 3.

WYKR. 3. SZCZEGÓLNE PRZYPADKI ZACHOWANIA SIĘ SZEREGÓW CZASOWYCH STÓP WZROSTU



U w a g a. *S* – odchylenie standardowe.
Ź r ó d ł o: opracowanie własne.

Jeśli przez ryzyko będziemy rozumieć odchylenia od stanu oczekiwanego, to z szeregiem A związane jest mniejsze ryzyko niż z szeregiem B (wykr. 3). Odchylenie standardowe dla szeregu A wynosi 1, a dla szeregu B przyjmuje wartość 3. Jednak inwestor przyzwyczaja się do takiej sytuacji i naturalne staje się dla niego, że z szeregiem A związane są mniejsze odchylenia, a z szeregiem B – większe. Z tego punktu widzenia ryzyko jest takie samo, ponieważ zmiany notowań są w pełni przewidywalne. Wartości wymiaru *MPP* wynoszą tutaj 2 i dla inwestora oznacza to, że po każdym odchyleniu stopy zwrotu w górę nastąpi takie samo co do siły odchylenie stopy zwrotu w dół.

Jeśli w szeregu A pojawią się zakłócenia (wykr. 3C), to inwestor stanie się już mniej pewny przyszłych zmian. Wartość odchylenia standardowego jest jednak ciągle mniejsza niż dla szeregu B, co wskazuje na mniejsze ryzyko. Tymczasem inwestor jest pewny zachowania szeregu B, a niepewny zachowania szeregu C. Obrazują to wartości wymiaru *MPP*. Szereg bez zakłóceń ma wymiar 2, a szereg z zakłóceniami w dążeniu do wartości średniej – 1,70.

Może się zdarzyć, że zakłócenia obniżą wartość odchylenia standardowego, jak w szeregu D w porównaniu do szeregu B (tutaj z 3 na 2,86). Jednak z punktu widzenia oczekiwań inwestora sytuacja jest już mniej stabilna, co obrazuje wymiar *MPP* = 1,89, a wartość mniejsza niż 2 wskazuje na obecność szumu.

Wymiar fraktalny służy tu jako metoda pozwalająca na ocenę stopnia stabilności wahań. Im wartości wymiaru fraktalnego bliższe 2, tym bardziej są stabilne, bez względu na zakres wahań. Niższe wartości sugerują pojawianie się zakłóceń: im bliżej górnej granicy odrzucenia hipotezy błądzenia losowego, tym więcej zakłóceń w szeregu stóp wzrostu.

STACJONARNOŚĆ SZEREGÓW CZASOWYCH

Weryfikacja hipotezy błądzenia losowego pozwala na lepsze poznanie właściwości statystycznych analizowanych szeregów czasowych cen instrumentów finansowych. Najczęściej weryfikuje się hipotezę o całkowitym braku korelacji w procesie generującym wartości szeregów czasowych oraz, rozszerzając pole badań, o braku jakichkolwiek zależności. Badanie braku zależności jest badaniem szerszym niż badanie braku korelacji. Wykluczenie zależności jednocześnie wyklucza występowanie korelacji, natomiast wykluczenie korelacji nie musi oznaczać jednoczesnego wykluczenia zależności. W przypadku braku korelacji jakiekolwiek liniowe modele cen historycznych nie pozwalają na lepsze przewidywanie cen w przyszłości. Jednak brak korelacji nie wyklucza występowania zależności nieliniowych, a takie mogłyby poprawić przewidywanie (Czekaj, Woś i Żarnowski, 2001). Dlatego badanie zależności jest przydatniejsze w praktyce. Proces błądzenia losowego opisuje się równaniem:

$$y_t = y_{t-1} + \varepsilon_t \tag{12}$$

gdzie:
y_t – szereg czasowy,
ε_t – proces niezależnych zmiennych losowych ciągłych o jednakowym rozkładzie ze średnią 0 i skończoną wariancją.

Szereg czasowy, który podlega procesowi błądzenia losowego, może dryfować w górę lub w dół tylko w rezultacie szoków losowych.

Z procesami stochastycznymi wiąże się pojęcie stacjonarności. Proces stochastyczny Y_t określa się jako stacjonarny, jeżeli spełnia jednocześnie trzy warunki:
1. $E(Y_t)$ = const – wartość średnia jest stała w czasie;
2. $Var(Y_t)$ = const – wariancja jest stała w czasie;
3. $Cov(Y_t,Y_{t+j}) = \sigma_j$ – wartość kowariancji dla dwóch momentów obserwacji zależy od odstępu między tymi momentami, a nie od ich wartości.

Jeżeli jeden z powyższych warunków nie jest spełniony, to taki proces określa się jako niestacjonarny.

Błądzenie losowe to przykład procesu niestacjonarnego, w którym wariancja jest zmienna w czasie (Maddala, 2006). Zmienia się także kowariancja pomiędzy sąsiednimi wartościami (Cryer, 1986). Wartości szeregu czasowego błądzenia losowego mogą znacznie odbiegać od wartości średniej.

Niestacjonarność szeregów czasowych stanowi poważny problem w przypadku modelowania zależności, ponieważ można otrzymywać dobre modele nawet dla zmiennych, dla których brak jest związku przyczynowo-skutkowego (Phillips, 1986). Dotyczy to zarówno trendu deterministycznego (Charemza i Deadman, 1997), jak i stochastycznego (Newbold i Davies, 1978). Analiza regresji powinna zostać przeprowadzona na stacjonarnych szeregach czasowych, aby była wiarygodna. Szeregi niestacjonarne można doprowadzić do stacjonarności poprzez różnicowanie ich wartości. Jednak nie każdy szereg można doprowadzić do stacjonarności, wyznaczając pierwsze różnice – niekiedy trzeba wyznaczać je więcej niż jeden raz. Szereg niestacjonarny, który można sprowadzić do szeregu stacjonarnego, obliczając przyrosty *d* razy, nazywa się szeregiem zintegrowanym stopnia *d* (Charemza i Deadman, 1997). Błądzenie losowe jest procesem zintegrowanym w stopniu pierwszym.

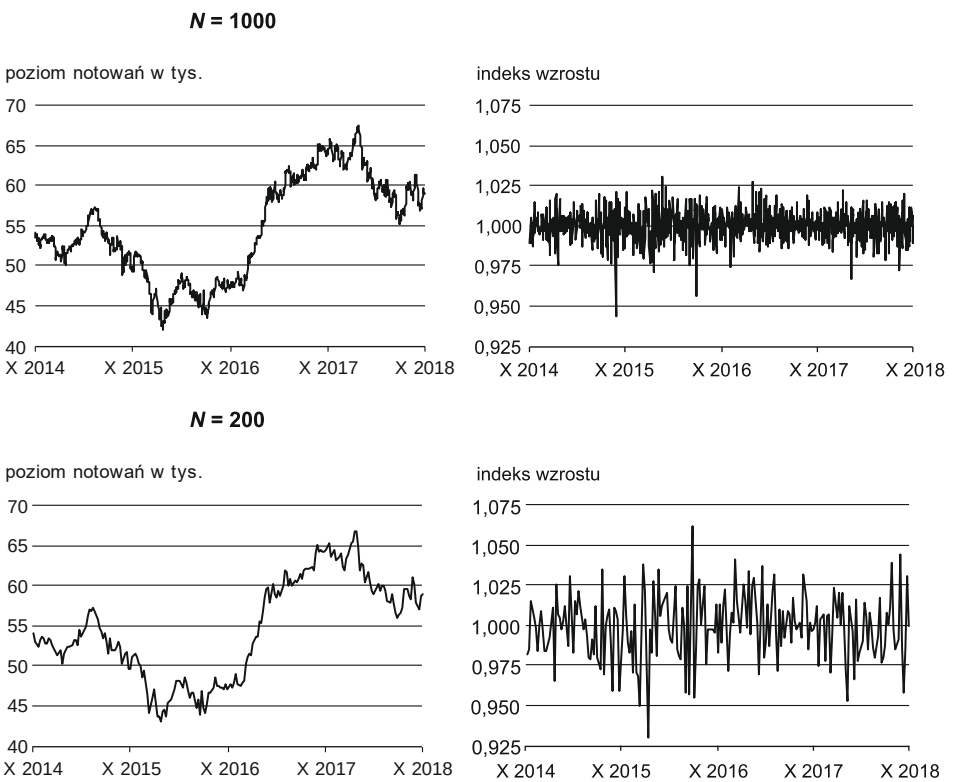
Istnieje wiele testów oceny stacjonarności szeregów czasowych. Do najpopularniejszych należą: DF, ADF, PP i KPSS (Dickey i Fuller, 1979, 1981; Kwiatkowski, Phillips, Schmidt i Shin, 1992; Phillips i Perron, 1988).

PRZYKŁADY ZASTOSOWANIA WYMIARU FRAKTALNEGO

Praktyczne zastosowanie wymiaru fraktalnego omówiono na przykładzie szeregu czasowego poziomów zamknięcia i stóp wzrostu indeksu giełdowego WIG oraz na przykładzie kształtowania się dziennych stóp wzrostu notowań wybranych spółek giełdowych.

Na wykr. 4 zaprezentowano kształtowanie się poziomów oraz indeksów wzrostu danych dziennych (*N* = 1000) i tygodniowych (*N* = 200) indeksu WIG w okresie od października 2014 r. do października 2018 r.

WYKR. 4. POZIOM NOTOWAŃ I INDEKSY WZROSTU INDEKSU GIEŁDOWEGO WIG



Ź r ó d ł o: opracowanie własne na podstawie danych GPW.

Uzyskane wartości wymiaru fraktalnego (tabl. 3) – dla *N* = 1000 jest to 1,3670, a dla *N* = 200 jest to 1,3716 – pozwalają na sklasyfikowanie szeregów jako procesów błądzenia losowego, a ściślej rzecz biorąc, uzyskane wartości nie pozwalają na odrzucenie hipotezy o błądzeniu losowym. Z kolei wartości wymiaru fraktalnego dla szeregów stóp wzrostu, odpowiednio *MPP* = 1,6916 oraz *MPP* = 1,6658, pozwalają na odrzucenie hipotezy błądzenia losowego na rzecz procesu o charakterze antypersystentnym. W szeregach stóp wzrostu występuje tendencja do oscylacji wokół poziomu przeciętnego.

TABL. 3. STATYSTYKI SZEREGÓW CZASOWYCH POZIOMÓW I PRZYROSTÓW INDEKSU GIEŁDOWEGO WIG

Szeregi a – poziomów b – przyrostów		*MPP*	Test ADF	
			t	*p*
N = 1000	a	1,3670	–1,7163	0,7436
	b	1,6916	–29,1944	0,0000
N = 200	a	1,3716	–1,7637	0,7188
	b	1,6658	–16,0766	0,0000

U w a g a. *t* – statystyka *t*-Studenta, *p* – poziom istotności.
Ź r ó d ł o: obliczenia własne.

Porównanie wyników dla szeregu *N* = 1000 z wynikami dla szeregu *N* = 200 prowadzi do wniosku, że częstotliwość obserwacji nie wpłynęła znacząco na uzyskane rezultaty. O ile w przypadku szeregów czasowych poziomów takie wyniki nie zaskakują, o tyle w przypadku szeregów stóp wzrostu są dość interesujące. Okazuje się, że wymiar fraktalny jest niewrażliwy na skalę wahań, ponieważ dla szeregu stóp wzrostu *N* = 1000 uzyskano odchylenie standardowe *S* = 0,0088, a dla szeregu *N* = 200 odchylenie standardowe jest wyraźnie większe i wynosi *S* = 0,0206. Tymczasem wartości wymiaru fraktalnego to odpowiednio *MPP* = 1,6916 oraz *MPP* = 1,6658. W świetle idei wymiaru fraktalnego takie wyniki są poprawne, ponieważ odchylenie standardowe mierzy skalę wahań, która dla szeregu stóp wzrostu *N* = 200 jest wyraźnie większa niż dla szeregu *N* = 1000, co pokazuje wykr. 4. Wymiar fraktalny mierzy zagęszczenie szeregu na płaszczyźnie – w przedmiotowym przykładzie, mimo różnej skali wahań, okazuje się ono zbliżone.

Wyniki testu ADF stanowią klasyczny rezultat testowania szeregów danych finansowych: szeregi poziomów okazują się niestacjonarne, a szeregi przyrostów (stóp wzrostu) – stacjonarne. Oznacza to, że nie można odrzucić hipotezy błądzenia losowego szeregów czasowych poziomów indeksu giełdowego. Ogólny obraz sytuacji, jaki uzyskano tutaj dzięki zastosowaniu wymiaru fraktalnego oraz testu ADF, jest podobny.

Analizę kształtowania szeregów czasowych dziennych stóp wzrostu wybranych spółek giełdowych przeprowadzono przy użyciu odchylenia standardowego oraz wymiaru fraktalnego (tabl. 4). Analizowano stopy wzrostu największych spółek giełdowych w latach 2015–2018. Pod uwagę brano spółki, których zakres dziennych zmian mieścił się w przedziale (–20%, 20%). Z tego powodu nie znalazła się wśród nich np. spółka JSW, której notowania w badanym okresie aż trzy razy zmieniały się powyżej 20% dziennie, co powoduje znaczne zawyżenie odchylenia standardowego w stosunku do pozostałych spółek.

Dla wybranej grupy spółek odchylenie standardowe dziennych stóp wzrostu kształtowało się w przedziale od 1,55 p.p. dla Asseco do 2,29 p.p. dla KGHM, a wymiaru fraktalnego – od 1,6439 dla Orange Polska do 1,7770 dla PKN Orlen.

TABL. 4. ODCHYLENIE STANDARDOWE I WYMIAR FRAKTALNY SZEREGÓW CZASOWYCH STÓP WZROSTU NOTOWAŃ WYBRANYCH SPÓŁEK W LATACH 2015–2018

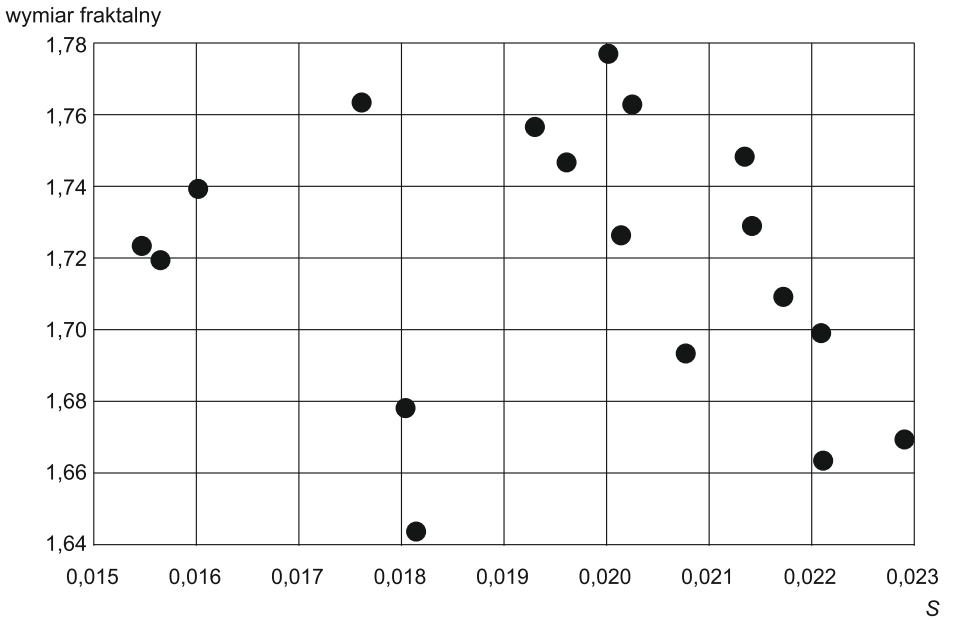
Spółki	*S*	Wymiar fraktalny	Wartość księgowa w tys. zł
Alior	0,0201	1,7265	6486
Asseco	0,0155	1,7234	5605
CCC	0,0221	1,6992	999
Cyfrowy Polsat	0,0176	1,7635	959
Energa	0,0214	1,7290	10263
Eurocash	0,0217	1,7093	975
KGHM	0,0229	1,6696	18524
Lotos	0,0196	1,7467	11955
LPP	0,0221	1,6637	2555
mBank	0,0208	1,6935	15214
Orange Polska	0,0181	1,6439	10501
PEKAO	0,0157	1,7195	22797
PGE	0,0202	1,7628	47196
PGNiG	0,0193	1,7566	35986
PKN Orlen	0,0200	1,7770	35634
PKO BP	0,0180	1,6783	37722
PZU	0,0160	1,7393	14169
Tauron	0,0213	1,7483	18968

Ź r ó d ł o: obliczenia własne.

Na wykr. 5 przedstawiono związek pomiędzy otrzymanymi wartościami odchylenia standardowego i wymiaru fraktalnego. Związek ten jest wprawdzie ujemny, lecz bardzo słaby. Wartość współczynnika korelacji liniowej Pearsona wynosi w tym przypadku –0,19.

Otrzymane wartości wymiaru fraktalnego pozwalają na sklasyfikowanie wszystkich analizowanych szeregów jako antypersystentnych. Jest to zgodne z naturą szeregów czasowych stóp wzrostu. Zważywszy na właściwości wymiaru fraktalnego oraz odchylenia standardowego, uzyskaną słabą zależność korelacyjną pomiędzy tymi miarami dla badanych szeregów czasowych należy uznać za oczekiwaną. Dowodzi ona, że miary te umożliwiają uchwycenie innych, unikalnych właściwości ocenianych szeregów czasowych. Zgodnie z uwagami metodologicznymi wymiar fraktalny uzupełnia odchylenie standardowe o informacje na temat stabilizacji zmian. Im wyższa jest wartość wymiaru fraktalnego, tym bardziej przewidywalna staje się zmienność.

WYKR. 5. ODCHYLENIE STANDARDOWE *S* A WYMIAR FRAKTALNY SZEREGÓW CZASOWYCH STÓP WZROSTU NOTOWAŃ



Ź r ó d ł o: jak przy wykr. 4.

W badaniach empirycznych dość często podnosi się znaczenie wielkości przedsiębiorstwa dla zmienności notowań. Z reguły notowania większych spółek odznaczają się większą płynnością, co sprzyja ograniczaniu ryzyka. W przedmiotowym badaniu związek wartości księgowej z odchyleniem standardowym jest ujemny, lecz bardzo słaby i statystycznie nieistotny ($r = –0{,}1019$; $p > 0{,}3$). Ujemny wynik oznacza, że przedsiębiorstwa o większej wartości księgowej charakteryzują się średnio mniejszą zmiennością notowań, ale jest to zależność znikoma. Tymczasem związek wartości księgowej z wymiarem fraktalnym jest dodatni i chociaż słaby, to statystycznie istotny ($p < 0{,}1$), a przy tym silniejszy od poprzedniego ($r = 0{,}3213$). Dodatni wynik oznacza, że przedsiębiorstwa o większej wartości księgowej charakteryzują się średnio większą wartością wymiaru fraktalnego, co sprzyja przewidywalności zmian. Wyniki te wymagają jednak szerszych badań.

PODSUMOWANIE

Wymiar fraktalny stanowi ciekawą alternatywę w stosunku do klasycznych metod oceny procesów kształtowania wartości szeregów czasowych. Do jego zalet należy dość prosta interpretacja: im mniejszy wymiar, tym silniejsze zjawi-

sko podtrzymania trendu, a im większy wymiar, tym silniejsze zjawisko powrotu do wartości średniej. Dodatkowym ułatwieniem interpretacyjnym jest unormowanie wymiaru fraktalnego w granicach [1, 2], tj. pomiędzy wymiarem euklidesowym prostej i płaszczyzny. Jak pokazano na przykładzie, możne on być w tej samej postaci stosowany do oceny zarówno poziomów szeregów czasowych, jak i ich przyrostów.

Interesujące jest porównanie wyników, jakie daje wymiar fraktalny dla szeregów stóp wzrostu, z wynikami odchylenia standardowego. Odchylenie standardowe pozwala na ocenę zmienności: im jest większe, tym większa zmienność. W klasycznej teorii rosnąca zmienność oznacza wyższe ryzyko, jednak większa zmienność nie musi się przekładać na wzrost nieprzewidywalności. Tymczasem idea wymiaru fraktalnego opiera się na zagęszczeniu szeregu na płaszczyźnie. Większy wymiar fraktalny to rosnące zagęszczenie, ale i wyższa przewidywalność, a także tendencja szeregu czasowego do powrotu do wartości średniej. Odchylenie standardowe pozwala tu na ocenę skali odchyleń od poziomu średniego. Pod tym względem można powiedzieć, że odchylenie standardowe i wymiar fraktalny się uzupełniają. Z drugiej strony niskie wartości wymiaru fraktalnego oznaczają również większą przewidywalność. Nieprzewidywalne natomiast stają się szeregi o wymiarze fraktalnym, który nie pozwala na odrzucenie hipotezy błądzenia losowego.

W interpretacji wymiaru fraktalnego należy rozróżnić analizę szeregów poziomów zmiennych i analizę szeregów stóp wzrostu. Szeregi poziomów będą miały z reguły niższe wymiary fraktalne – bliższe 1. Im wartość wymiaru fraktalnego bliższa 1, tym silniejszy trend w ocenianym zjawisku. Wyższe wartości wymiaru fraktalnego będą upodabniały szeregi poziomów do szeregów błądzenia losowego. Inaczej przedstawia się sytuacja w przypadku szeregów stóp wzrostu. Tutaj wartości wymiaru fraktalnego są z reguły wyższe. Im wartości wymiaru fraktalnego bliższe 2, tym bardziej są stabilne, bez względu na zakres wahań. Natomiast niższe wartości wymiaru fraktalnego sugerują pojawianie się zakłóceń: im niższa wartość wymiaru fraktalnego szeregów stóp wzrostu, tym więcej zakłóceń w takim szeregu. Wymiar fraktalny stanowi tu metodę pozwalającą na ocenę stopnia stabilności wahań.

Ogólny wynik klasyfikacyjny uzyskany w analizowanym przykładzie jest zgodny z wynikami testów stacjonarności ADF. Jednak w porównaniu do testów ADF unormowanie wymiaru fraktalnego oraz jego intuicyjna interpretacja dają nowe możliwości oceny kształtowania wartości szeregów czasowych. Pewnym ograniczeniem stosowania wymiaru fraktalnego jest konieczność badania stosunkowo długich szeregów czasowych, umożliwiających dokonywanie podziałów, chociaż – jak pokazano na przykładzie – przejście z danych dziennych na dane tygodniowe, a tym samym skrócenie długości szeregu z 1000 do 200 obserwacji, nie spowodowało istotnych zmian wyników.

BIBLIOGRAFIA

Bhatt, S. J., Dedania, H. V., Shah, V. R. (2015). Fractal dimensional analysis in financial time series. *International Journal of Financial Management*, *5*(2), 57–62.

Buła, R. (2017). Analiza wymiaru fraktalnego spółek notowanych na Giełdzie Papierów Wartościowych w Warszawie – aspekty metodyczne. *Nauki o Finansach*, *1*(30), 9–27.

Cervantes-De la Torre, F., González-Trejo, J. I., Real-Ramírez, C. A., Hoyos-Reyes, L. F. (2013). Fractal dimension algorithms and their application to time series associated with natural phenomena. *Journal of Physics*: *Conference Series*, *475*, 1–10. DOI: 10.1088/1742-6596/475/1/012002.

Charemza, W., Deadman, D. (1997). *Nowa ekonometria*. Warszawa: Polskie Wydawnictwo Ekonomiczne.

Chen, Y. (2013). A Set of Formulae on Fractal Dimension Relations and its Application to Urban Form. *Chaos, Solitons & Fractals*, *54*, 150–158. DOI: 10.1016/j.chaos.2013.07.010.

Cryer, J. D. (1986). *Time Series Analysis*. Boston: Duxbury Press.

Czekaj, J., Woś, M., Żarnowski, J. (2001). *Efektywność giełdowego rynku akcji w Polsce: z perspektywy dziesięciolecia*. Warszawa: Wydawnictwo Naukowe PWN.

Dickey, D. A., Fuller, W. A. (1979). Distributions of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, *74*(366), 427–431. DOI: 10.1080/01621459.1979.10482531.

Dickey, D. A., Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, *49*(4), 1057–1072. DOI: 10.2307/1912517.

Dubuc, B., Quiniou, J. F., Roques-Carmes, C., Tricot, C., Zucker, S. W. (1989). Evaulating the Fractal Dimension of Profiles. *Physical Review A*, *39*(3), 1500–1512. DOI: 10.1103/PhysRevA.39.1500.

Gómez, C., Mediavilla, Á., Hornero, R., Abásolo, D., Fernández, A. (2009). Use of the Higuchi's fractal dimension for the analysis of MEG recordings from Alzheimer's disease patients. *Medical Engineering & Physics*, *31*(3), 306–313. DOI: 10.1016/j.medengphy.2008.06.010.

Guoqiang, S. (2002). Fractal dimension and fractal growth of urbanized areas. *International Journal of Geographical Information Science*, *16*(5), 419–437. DOI: 10.1080/13658810210137013.

Halley, J. M., Kunin, W. E. (1999). Extinction Risk and the 1/f Family of Noise Models. *Theoretical Population Biology*, *56*(3), 215–230. DOI: 10.1006/tpbi.1999.1424.

Harne, B. P. (2014). Higuchi Fractal Dimension Analysis of EEG Signal Before and After OM Chanting to Observe Overall Effect on Brain. *International Journal of Electrical and Computer Engineering*, *4*(4), 585–592. DOI: 10.11591/ijece.v4i4.5800.

Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of American Society of Civil Engineers*, *116*, 770–799.

Kale, M., Butar Butar, F. (2011). Fractal analysis of time series and distribution properties of Hurst exponent. *Journal of Mathematical Sciences & Mathematics Education*, *5*(1), 8–19.

Kapecka, A. (2013). Fractal Analysis of Financial Time Series Using Fractal Dimension and Pointwise Hölder Exponents. *Dynamic Econometric Models*, (13), 107–125. DOI: 10.12775/DEM.2013.006.

Kwiatkowski, D., Phillips, P., Schmidt, P., Shin, Y. (1992). Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We that Economic Time Series Have a Unit Root?. *Journal of Econometrics*, *54*(1–3), 159–178.

Maddala, G. S. (2006). *Ekonometria*. Warszawa: Wydawnictwo Naukowe PWN.

Mandelbrot, B. B. (1982). *The Fractal Geometry of Nature*. New York: W.H. Freeman and Company.

Mosdorf, R. (1997). *Dynamiczny model wrzenia na podstawie metody chaosu deterministycznego*. Białystok: Wydawnictwo Politechniki Białostockiej.

Newbold, P., Davies, N. (1978). Error Mis-Specification and Spurious Regressions. *International Economic Review*, *19*(2), 513–519. DOI: 10.2307/2526317.

Peters, E. E. (1997). *Teoria chaosu a rynki kapitałowe*. Warszawa: WIG-Press.

Phillips, P. C. B. (1986). Understanding spurious regressions in econometrics. *Journal of Econometrics*, *33*(3), 311–340. DOI: 10.1016/0304-4076(86)90001-1.

Phillips, P. C. B., Perron, P. (1988). Testing for a Unit Root in Time Series Regression. *Biometrika,* (*75*), 335–346.

Przekota, G., Przekota, D. (2004). Szacowanie wymiaru fraktalnego szeregów czasowych kursów walut metodą podziału pola. *Badania Operacyjne i Decyzje*, *14*(3–4), 67–82.

Siemieniuk, N. (2001). *Fraktalne właściwości polskiego rynku kapitałowego*. Białystok: Wydawnictwo Uniwersytetu w Białymstoku.

Sy-Sang, L., Feng-Yuan, C. (2009). Fractal dimensions of time sequences. *Physica A: Statistical Mechanics and its Applications*, *388*(15), 3100–3106. DOI: 10.1016/j.physa.2009.04.011.

Zawadzki, H. (1996). *Chaotyczne systemy dynamiczne: elementy teorii i wybrane przykłady ekonomiczne*. Katowice: Akademia Ekonomiczna.

Zeug-Żebro, K. (2015). Zastosowanie wybranych metod szacowania wymiaru fraktalnego do oceny poziomu ryzyka finansowych szeregów czasowych. *Studia Ekonomiczne. Zeszyty Naukowe*, *227*, 109–124.

Zwolankowska, M. (2001). *Fraktalna geometria polskiego rynku akcji*. Szczecin: Wydawnictwo Naukowe Uniwersytetu Szczecińskiego.

Wiadomości Statystyczne. The Polish Statistician, 2019, vol. 64, 9, 25—36 DOI: 10.5604/01.3001.0013.7593
Statystyka w praktyce / Statistics in practice

Zmiana sektorowej struktury pracujących w przekroju województw[1]

Katarzyna Wawrzyniak[a], **Barbara Batóg**[b]

Streszczenie. Celem badania jest identyfikacja prawidłowości w zakresie zmian w strukturze pracujących według sektorów ekonomicznych w przekroju województw. Analizie poddano lata 2010–2016. Wykorzystano dane Głównego Urzędu Statystycznego o pracujących według faktycznego miejsca pracy w sześciu sektorach ekonomicznych (grupach sekcji). Określono stopień podobieństwa sektorowej struktury pracujących w województwach do struktury w Polsce. Zróżnicowanie udziału procentowego pracujących według sektorów ekonomicznych w poszczególnych województwach oszacowano za pomocą względnego wskaźnika Herfindahla-Hirschmana. Zbadano również podobieństwo sektorowej struktury pracujących w województwach w ujęciu łańcuchowym (rok do roku), wykorzystując miarę podobieństwa struktur. Województwa pogrupowano według stopnia podobieństwa struktur, co pozwoliło wskazać regiony charakteryzujące się największymi i najmniejszymi zmianami w analizowanym okresie.

Z badania wynika, że w latach 2010–2016 zmiany w strukturze pracujących według sektorów ekonomicznych w poszczególnych województwach zachodziły powoli, przy czym w większości województw spadł udział pracujących w rolnictwie, a wzrósł w usługach związanych z działalnością finansową i ubezpieczeniową oraz obsługą rynku nieruchomości. Największe zróżnicowanie udziału procentowego w strukturze pracujących zaobserwowano w woj. lubelskim, a najmniejsze – w małopolskim. Zmiany w strukturze pracujących według sektorów były najbardziej dynamiczne na początku badanego okresu (2011/2010), natomiast pod koniec – coraz wolniejsze.

Słowa kluczowe: województwa, sektory ekonomiczne, struktura pracujących, wskaźnik Herfindahla-Hirschmana, miara podobieństwa struktur

The change in the sectoral structure of employed persons by voivodships

Summary. The aim of this paper is to identify the regularities in the pattern of changes in the employment structure by economic sectors in Polish voivodships in the period of 2010–2016. The research uses the data on persons employed by the actual workplace in six economic sectors (groups of sections), obtained from Statistics Poland. The research determines the degree of similarity between the sectoral structure of employment of each examined voivodship and such structure of whole Poland. The diversity of percentages of persons employed in particular economic sectors within each examined voivodship was determined using the Herfindahl-Hirschman relative index. The research also examines the similarity of sectoral structures of employment in voivodships in the chain ap-proach (year-on-year), using the similarity measure of structures. The voivodships were grouped according to the degree of similarity of structures, which, in turn, made it possible to identify the voivodships where the changes observed in the analysed period were most and least significant.

The research demonstrates that in the years 2010–2016, the changes in the structure of employment by economic sectors in the examined voivodships took place relatively slowly, and that in most voivodships, the employment rate in agriculture decreased, while in services for financial, insurance and real estate sectors, it increased. The highest diversification in the structure of employment was observed in Lubelskie voivodship, and the lowest in Małopolskie voivodship. The most dynamic changes in the sectoral structure of employment were observed at the beginning of the analysed period (2011/2010), whereas towards its end, these changes became slower.

Keywords: voivodships, economic sectors, structure of employment, Herfindahl-Hirschman index, similarity measure of structures

JEL: J21, R23, C10

[1] Artykuł został opracowany na podstawie referatu wygłoszonego na II Kongresie Statystyki Polskiej, który odbył się w Warszawie w dniach 10–12 lipca 2018 r.

[a] Zachodniopomorski Uniwersytet Technologiczny w Szczecinie, Wydział Ekonomiczny.

[b] Uniwersytet Szczeciński, Wydział Nauk Ekonomicznych i Zarządzania.

Celem badania opisanego w niniejszym artykule jest identyfikacja prawidłowości w zakresie zmian struktury pracujących według sektorów ekonomicznych w przekroju województw. Badanie obejmowało lata 2010–2016 i składało się z trzech etapów. Najpierw sprawdzono, jaki jest stopień podobieństwa struktury pracujących według sektorów ekonomicznych w poszczególnych województwach w odniesieniu do struktury w Polsce. Następnie oceniono zróżnicowanie udziału procentowego pracujących według sektorów ekonomicznych w województwach w analizowanym okresie, po czym zbadano stopień podobieństwa struktury pracujących według sektorów ekonomicznych w województwach w ujęciu łańcuchowym (rok do roku).

Wykorzystano dane Głównego Urzędu Statystycznego (GUS) o pracujących według faktycznego miejsca pracy i rodzaju działalności (bez pracujących w jednostkach budżetowych działających w zakresie obrony narodowej i bezpieczeństwa) w sześciu sektorach ekonomicznych (grupach sekcji) w województwach w Polsce w latach 2010–2016. Poszczególne sektory ekonomiczne zostały zdefiniowane następująco:

- sektor 1: sekcja A – Rolnictwo, leśnictwo, łowiectwo i rybactwo;
- sektor 2: sekcja B – Górnictwo i wydobywanie, sekcja C – Przetwórstwo przemysłowe, sekcja D – Wytwarzanie i zaopatrywanie w energię elektryczną, gaz, parę wodną, gorącą wodę i powietrze do układów klimatyzacyjnych;
- sektor 3: sekcja F – Budownictwo;
- sektor 4: sekcja G – Handel hurtowy i detaliczny; naprawa pojazdów samochodowych, włączając motocykle, sekcja H – Transport i gospodarka magazynowa, sekcja I – Zakwaterowanie i gastronomia, sekcja J – Informacja i komunikacja;
- sektor 5: sekcja K – Działalność finansowa i ubezpieczeniowa, sekcja L – Działalność związana z obsługą rynku nieruchomości;
- sektor 6: sekcja M – Działalność profesjonalna, naukowa i techniczna, sekcja N – Działalność w zakresie usług administrowania i działalność wspierająca, sekcja O – Administracja publiczna i obrona narodowa; obowiązkowe zabezpieczenia społeczne, sekcja P – Edukacja, sekcja Q – Opieka zdrowotna i pomoc społeczna, sekcja R – Działalność związana z kulturą, rozrywką i rekreacją, sekcja S – Pozostała działalność usługowa.

METODA BADANIA

Badanie zmian w strukturze sektorowej pracujących w gospodarce narodowej stanowi istotny element oceny stopnia rozwoju gospodarki danego kraju. Podstawą tej oceny może być teoria trzech sektorów[2], zgodnie z którą kraje znajdu-

[2] Teoria trzech sektorów: rolniczego, przemysłowego i usługowego została stworzona przez Allana G. B. Fishera, Colina Clarka oraz Jeana Fourastiégo w latach 30. XX w. W literaturze polskiej zwartą publikacją dotyczącą trzech sektorów gospodarki jest m.in. praca Kwiatkowskiego (1980).

jące się na wyższym poziomie rozwoju charakteryzują się relatywnie wysokim udziałem sektora usługowego, umiarkowanym udziałem sektora przemysłowego i bardzo niskim udziałem rolnictwa (3–5%) w strukturze zatrudnienia (Zajdel, 2010). Znając strukturę sektorową pracujących, można stwierdzić, na jakim etapie rozwoju dany kraj znajduje się w danym momencie, natomiast porównanie tych struktur w czasie i przestrzeni daje przesłanki do stwierdzenia, czy rozwój kraju zmierza we właściwym kierunku oraz jaka jest jego sytuacja w stosunku do innych krajów. Tego typu badania i analizy prezentowane są w licznych publikacjach dotyczących struktury pracujących zarówno w skali kraju (np. Batóg i Batóg, 2001; Dworak i Malarska, 2010; Zajdel, 2010), jak i regionów/województw (np. Adamczyk, 2012; Dąbrowska, 2013; Kosmalski, 2008; Walkowiak, 2016), przy czym rozpatrywana jest w nich struktura trójsektorowa, a także struktura pracujących według sekcji PKD. Porównanie zmian w strukturze zatrudnienia w Polsce na tle krajów Unii Europejskiej (UE) było przedmiotem rozważań Maliny (2006, 2008), która dokonała również pogrupowania krajów UE pod względem podobieństwa struktury zatrudnienia w wybranych latach z wykorzystaniem metody Warda. Zmiany zachodzące w trójsektorowej strukturze pracujących w Polsce na tle krajów UE na przestrzeni wybranych lat zostały opisane m.in. przez Klembowską (2008), Klimczyka (2008) oraz Kwiatkowską (2016). Wymienieni autorzy na podstawie przeprowadzonych badań zauważyli, że odsetek pracujących w rolnictwie w Polsce zmniejszał się systematycznie, ale zawsze był wyższy niż w najbardziej rozwiniętych krajach UE.

Do zbadania zmian w strukturze pracujących według sektorów ekonomicznych w ujęciu regionalnym[3] w Polsce zastosowano trzy miary:

- miarę podobieństwa (odległości) struktur, opartą na metryce miejskiej, obliczaną według wzoru:

$$v_r = \frac{\sum_{i=1}^{k} |\alpha_{ir} - \alpha_i|}{2} \tag{1}$$

- względny wskaźnik Herfindahla-Hirschmana (*HH*), obliczany ze wzoru (Suchecki, 2010):

$$HH_t = \frac{\sum_{i=1}^{k} \alpha_{it}^2 - \frac{1}{k}}{1 - \frac{1}{k}} \tag{2}$$

- miarę podobieństwa struktur, obliczaną według wzoru (Wędrowska, 2012):

$$v_t = \frac{\sum_{i=1}^{k} |\alpha_{it} - \alpha_{it-1}|}{2} \tag{3}$$

[3] W artykule pojęcia *region* i *województwo* są używane zamiennie.

gdzie:
k – liczba składowych w danej strukturze,
α_i – udział i-tej składowej w danej strukturze w Polsce,
α_{ir} – udział i-tej składowej w danej strukturze w województwie r,
α_{it} – udział i-tej składowej w danej strukturze w okresie t,
α_{it-1} – udział i-tej składowej w danej strukturze w okresie $t - 1$.

Pierwszą miarę wykorzystano do zbadania podobieństwa struktury pracujących według sektorów ekonomicznych w województwach do struktury w Polsce, drugą – do oceny stopnia zróżnicowania udziału procentowego pracujących według sektorów ekonomicznych w Polsce i w poszczególnych województwach, natomiast trzecią – do zbadania podobieństwa struktury pracujących w Polsce i w województwach w ujęciu łańcuchowym (rok do roku).

ZMIANY W STRUKTURZE PRACUJĄCYCH W SEKTORACH

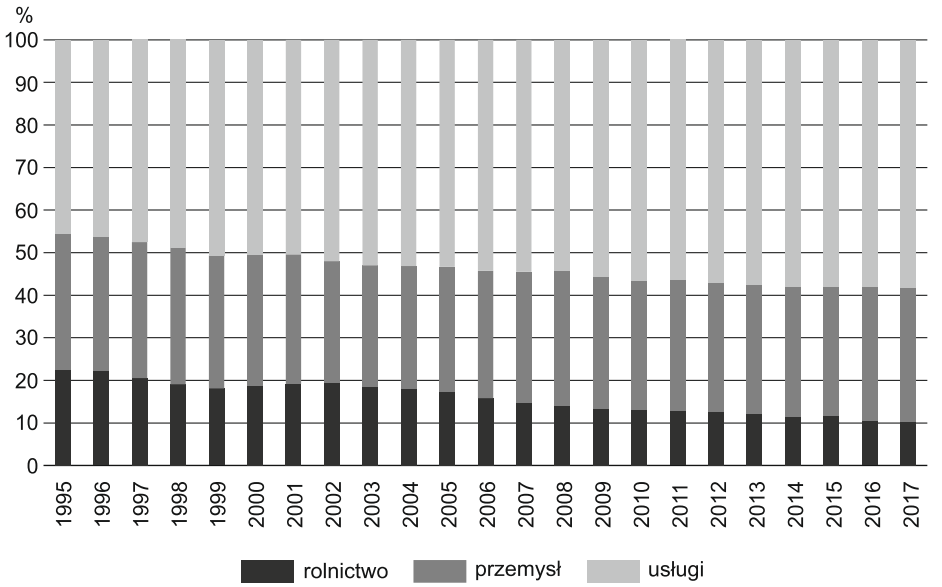
Badanie zmian w strukturze sektorowej pracujących w ujęciu regionalnym rozpoczęto od sprawdzenia, w jakim stopniu struktura pracujących w Polsce – z uwzględnieniem podziału na trzy sektory: rolnictwo, przemysł i usługi – dążyła do struktury zgodnej z teorią trzech sektorów. Z danych pochodzących z Badania Aktywności Ekonomicznej Ludności (BAEL), przedstawionych na wykr. 1, wynika, że w latach 1995–2017 udział pracujących w rolnictwie spadł z 22,6% do 10,2%, udział pracujących w usługach wzrósł z 45,4% do 57,9%, natomiast udział pracujących w przemyśle w całym okresie utrzymywał się na poziomie od 30% do 32%. Zaobserwowane zmiany świadczą o tym, że Polskę – z punktu widzenia tego kryterium – można obecnie zaliczyć do krajów o wyższym poziomie rozwoju gospodarczego.

Następnie analizie poddano strukturę pracujących w Polsce według sześciu sektorów ekonomicznych w latach 2010–2016. Pozwoliło to zweryfikować prawidłowości w zakresie zmian w strukturze pracujących na niższym poziomie agregacji, czyli z uwzględnieniem podziału sektora przemysłowego na sektory 2 i 3 oraz sektora usługowego na sektory 4, 5 i 6. Jedynie w przypadku rolnictwa nie dokonano dezagregacji danych o pracujących[4]. Na podstawie wykr. 2 można stwierdzić, że w badanym okresie analizowana struktura pracujących w Polsce nie uległa istotnym zmianom oraz że widoczne są następujące prawidłowości:

- udział pracujących w rolnictwie (sektor 1) zmniejszył się z 17,3% do 16,0%;
- najbardziej wzrósł udział pracujących w sektorze 6 – od 27,0% do 28,1%;
- w pozostałych czterech sektorach udział pracujących był prawie jednakowy w całym badanym okresie i wynosił: w sektorze 2 – ok. 21%, w sektorze 3 – ok. 6%, w sektorze 4 – ok. 25%, w sektorze 5 – ok. 4%.

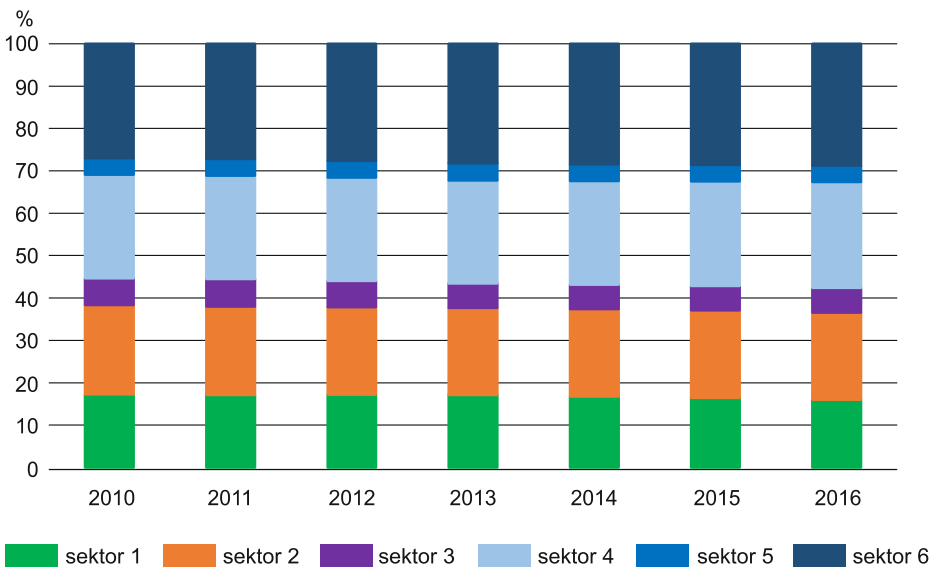
[4] Należy zaznaczyć, że dane o pracujących w sześciu sektorach ekonomicznych pochodzą z Banku Danych Lokalnych (BDL) GUS. Ze względu na odmienny sposób gromadzenia danych udział procentowy pracujących w rolnictwie, przemyśle i usługach różni się od udziału wyznaczonego na podstawie badania BAEL, różnice te nie są jednak duże.

WYKR. 1. TRÓJSEKTOROWA STRUKTURA PRACUJĄCYCH



Ź r ó d ł o: BAEL GUS.

WYKR. 2. STRUKTURA PRACUJĄCYCH WEDŁUG SEKTORÓW EKONOMICZNYCH



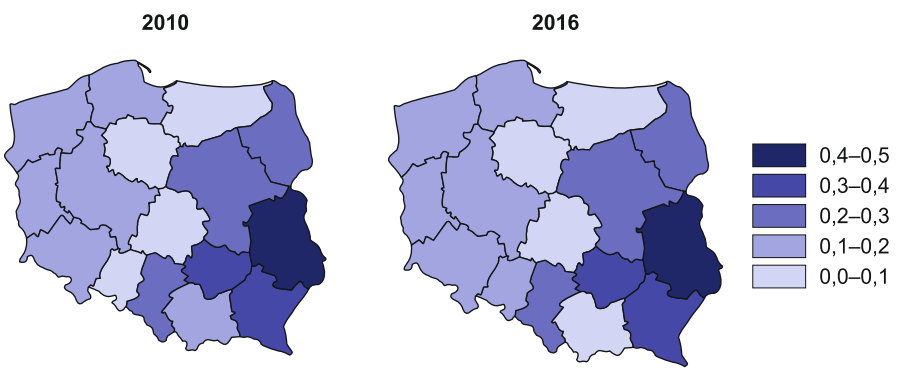
Ź r ó d ł o: BDL GUS.

PODOBIEŃSTWO STRUKTURY PRACUJĄCYCH WEDŁUG SEKTORÓW EKONOMICZNYCH W WOJEWÓDZTWACH

Nieznaczne zmiany, jakie zaobserwowano w strukturze pracujących według sektorów ekonomicznych w Polsce, skłoniły do sprawdzenia tej prawidłowości w ujęciu regionalnym. W tym celu w poszczególnych latach obliczono miarę podobieństwa struktury pracujących według sektorów ekonomicznych w województwach do struktury w Polsce (wzór 1).

W całym analizowanym okresie wartości miary wskazały na niewielkie zmiany w podobieństwie struktury pracujących w poszczególnych województwach do struktury w Polsce. Weryfikacja zmian strukturalnych była możliwa dopiero w okresie siedmiu lat (mapa 1 przedstawia odległości struktur sektorowych). Natomiast na wykr. 3 zaprezentowano strukturę sektorową Polski oraz województw kujawsko-pomorskiego (największe podobieństwo) i lubelskiego (najmniejsze podobieństwo).

MAPA 1. ODLEGŁOŚCI STRUKTURY SEKTOROWEJ WOJEWÓDZTW DO STRUKTURY SEKTOROWEJ POLSKI



U w a g a. Przedziały są domknięte prawostronnie.
Ź r ó d ł o: opracowanie własne na podstawie BDL GUS.

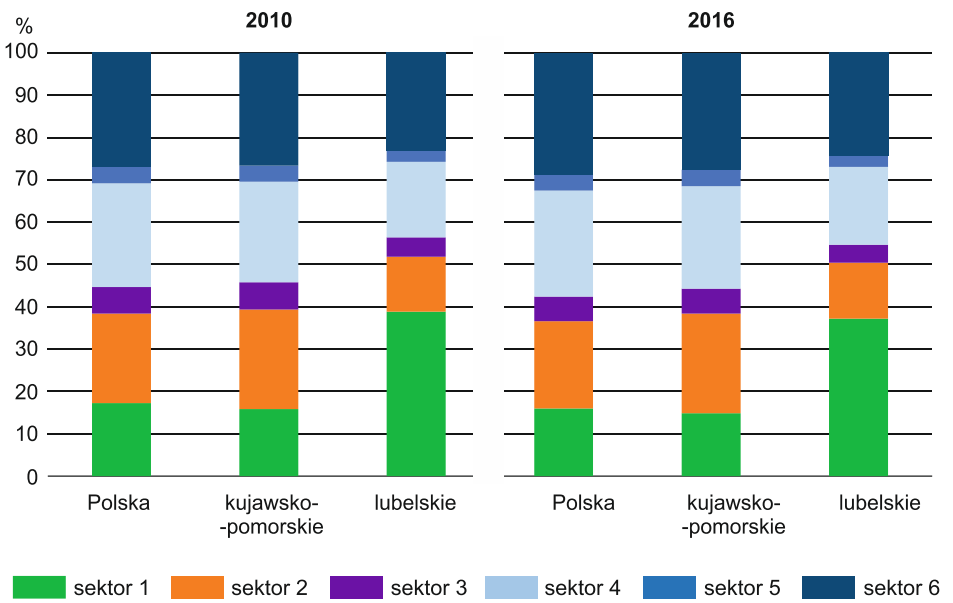
Badanie podobieństwa sektorowej struktury pracujących w poszczególnych województwach do struktury sektorowej w Polsce umożliwiło zaobserwowanie następujących prawidłowości (mapa 1, wykr. 3):

- we wszystkich badanych latach miara podobieństwa sektorowej struktury pracujących w województwach w odniesieniu do Polski kształtowała się na poziomie od 0,05 do 0,45, a więc zróżnicowanie stopnia podobieństwa struktury pracujących w tych latach było dość duże;
- zarówno w 2010 r., jak i w 2016 r. najbardziej podobną sektorową strukturę pracujących do struktury w Polsce miały województwa: kujawsko-pomorskie,

warmińsko-mazurskie i łódzkie, natomiast najmniej podobną – woj. lubelskie (wyraźna przewaga pracujących w sektorze 1);

- podobieństwo sektorowej struktury pracujących w województwach do struktury w Polsce pozwoliło wyróżnić po pięć grup województw zarówno w 2010 r., jak i w 2016 r., przy czym jedynie dwa województwa przesunęły się do innej grupy:
  – w przypadku woj. opolskiego zmniejszyło się podobieństwo struktury pracujących do struktury w Polsce, gdyż zwiększyły się różnice między udziałem procentowym w sektorach 3 i 4,
  – w przypadku woj. małopolskiego zwiększyło się podobieństwo struktury pracujących do struktury w Polsce, czyli zmniejszyły się różnice między udziałem w poszczególnych sektorach.

WYKR. 3. STRUKTURA SEKTOROWA POLSKI ORAZ WOJEWÓDZTW KUJAWSKO-POMORSKIEGO I LUBELSKIEGO



Ź r ó d ł o: jak przy mapie 1.

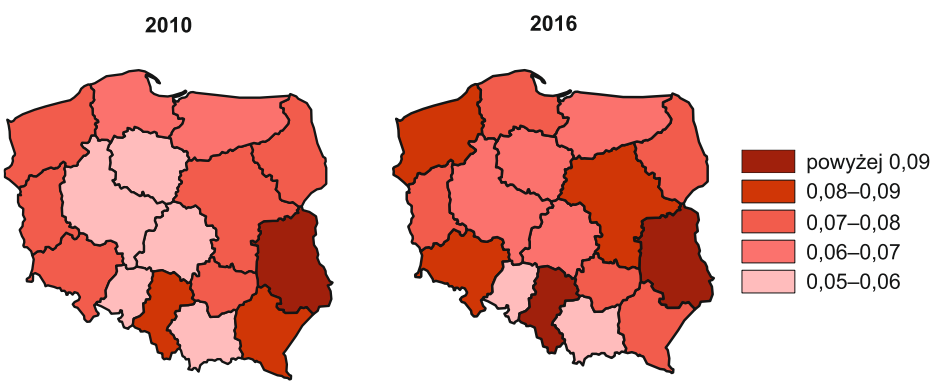
ZRÓŻNICOWANIE UDZIAŁU PRACUJĄCYCH WEDŁUG SEKTORÓW EKONOMICZNYCH W WOJEWÓDZTWACH

Ze względu na duże podobieństwo struktur pracujących w województwach w odniesieniu do Polski zdecydowano się na ocenę stopnia zróżnicowania udziału procentowego pracujących według sektorów ekonomicznych w Polsce

i w województwach w badanym okresie. Wykorzystano w tym celu względny wskaźnik *HH* (wzór 2), który przyjmuje wartości z przedziału [0, 1). Wartość 0 otrzymuje się w przypadku, gdy udział wszystkich składowych w danej strukturze jest jednakowy. Im wartość wskaźnika bliższa 0, tym mniejsze zróżnicowanie udziału składowych w badanej strukturze. Wartość wskaźnika bliska 1 oznacza, że udział jednej składowej jest bardzo wysoki (prawie 1), a pozostałych składowych – bardzo niski (prawie 0). W ten sposób wyodrębniono województwa i sektory charakteryzujące się największymi i najmniejszymi zmianami udziału pracujących.

Względny wskaźnik *HH* dla Polski wzrósł od 0,057 w 2010 r. do 0,062 w 2016 r. (mapa 2). Jego niewielki wzrost w ciągu siedmiu lat jest konsekwencją zmian tylko w dwóch sektorach – spadku udziału pracujących w sektorze 1 (rolnictwo) i wzrostu w sektorze 6.

MAPA 2. WARTOŚCI WSKAŹNIKA *HH* DLA WOJEWÓDZTW



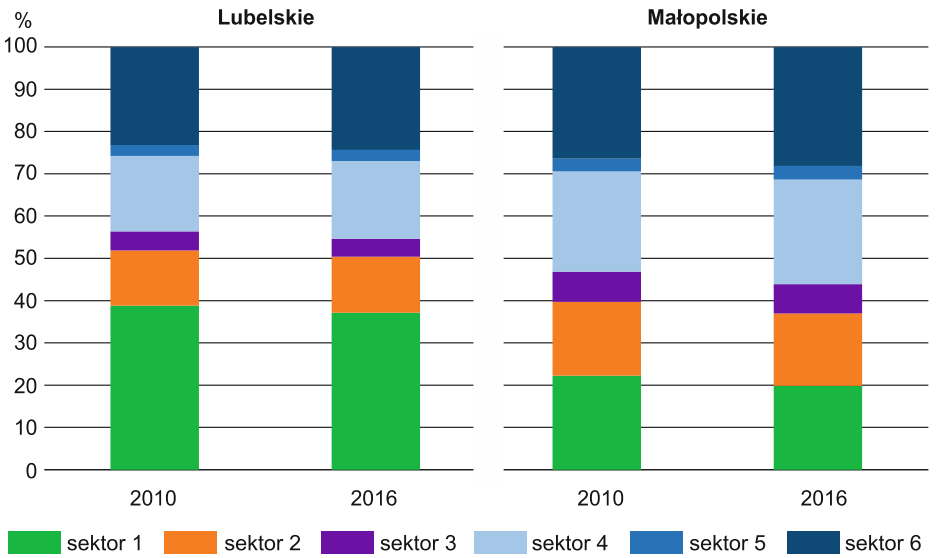
U w a g a. Jak przy mapie 1.
Ź r ó d ł o: jak przy mapie 1.

Badanie stopnia zróżnicowania udziału pracujących według sześciu sektorów w województwach z wykorzystaniem wskaźnika *HH* pozwoliło zaobserwować, że:

- zróżnicowanie udziału procentowego pracujących w poszczególnych województwach nie jest duże (świadczą o tym niskie wartości wskaźnika *HH*);
- w 2016 r. w porównaniu do 2010 r. w większości województw zróżnicowanie udziału pracujących się zwiększyło (wzrost wartości wskaźnika *HH*) i jedynie w dwóch – opolskim i małopolskim – pozostało bez zmian;
- największym zróżnicowaniem zarówno w 2010 r., jak i w 2016 r. charakteryzowało się woj. lubelskie (znaczny udział pracujących w rolnictwie), a naj-

mniejszym – woj. małopolskie (prawie identyczny udział pracujących w sektorach 1, 4 i 6); odmienność struktury tych regionów prezentuje wykr. 4.

WYKR. 4. STRUKTURA SEKTOROWA WOJEWÓDZTW O NAJMNIEJSZEJ (małopolskie) I NAJWIĘKSZEJ (lubelskie) WARTOŚCI WSKAŹNIKA HH



Ź r ó d ł o: jak przy mapie 1.

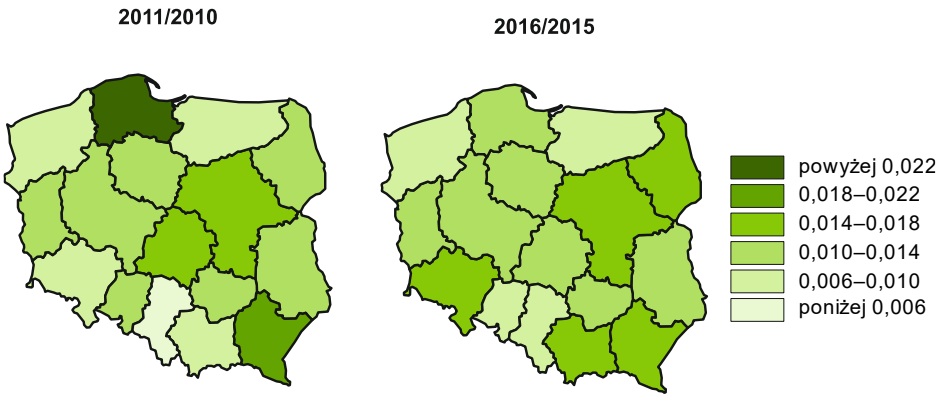
PODOBIEŃSTWO STRUKTURY PRACUJĄCYCH WEDŁUG SEKTORÓW EKONOMICZNYCH W WOJEWÓDZTWACH W UJĘCIU ŁAŃCUCHOWYM

Wartości wskaźnika *HH* w analizowanych latach były dość niskie, ale wykazywały zmiany w czasie. W związku z tym zdecydowano się na zbadanie podobieństwa struktury pracujących w Polsce i w województwach w ujęciu łańcuchowym (rok do roku), wykorzystując miarę podobieństwa obliczoną zgodnie ze wzorem (3).

W całym badanym okresie miara podobieństwa struktury dla Polski rok do roku kształtowała się na poziomie ok. 0,01. Niska wartość tej miary świadczy o bardzo powolnych zmianach w strukturze sektorowej pracujących w latach 2010–2016.

Na mapie 3 przedstawiono odległości struktur sektorowych województw dla lat 2011/2010 oraz 2016/2015. Na wykr. 5 zaprezentowano strukturę sektorową w woj. pomorskim w latach 2010 i 2011 oraz 2015 i 2016 jako przykład struktury o największym zróżnicowaniu udziału pracujących.

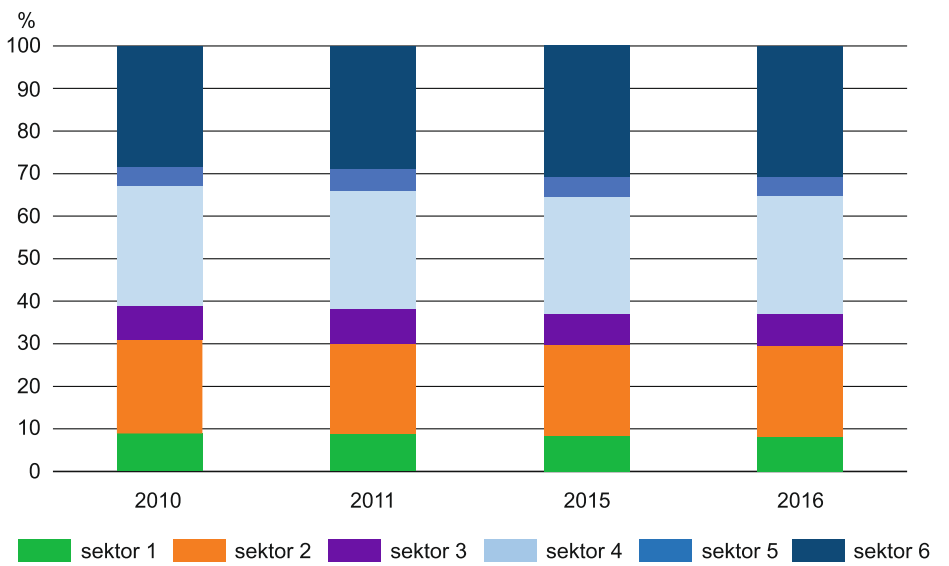
MAPA 3. ODLEGŁOŚCI STRUKTUR SEKTOROWYCH WOJEWÓDZTW



U w a g a. Jak przy mapie 1.
Ź r ó d ł o: jak przy mapie 1.

Badanie podobieństwa sektorowej struktury pracujących w poszczególnych województwach z wykorzystaniem metryki miejskiej w ujęciu łańcuchowym umożliwiło zaobserwowanie następujących prawidłowości (mapa 3, wykr. 5):

WYKR. 5. STRUKTURA SEKTOROWA W WOJEWÓDZTWIE POMORSKIM



Ź r ó d ł o: jak przy mapie 1.

- w latach 2010–2016 sektorowa struktura pracujących w województwach zmieniała się z roku na rok bardzo nieznacznie (o czym świadczą bardzo niskie wartości miary podobieństwa);
- w 2011 r. w stosunku do 2010 r. największe zmiany w strukturze pracujących miały miejsce w woj. pomorskim (w sektorach 2 i 5), a najmniejsze – w śląskim; poziom metryki miejskiej dla tych lat umożliwił wydzielenie sześciu grup województw;
- w 2016 r. w stosunku do poprzedniego roku dynamika zmian w sektorowej strukturze pracujących uzasadniała wyróżnienie czterech grup województw, gdyż zróżnicowanie dynamiki zmian było znacznie mniejsze – w większości województw zmiany te były znacznie wolniejsze niż na początku badanego okresu;
- jedynie w woj. mazowieckim dynamika zmian w sektorowej strukturze pracujących była zbliżona zarówno na początku, jak i na końcu badanego okresu.

PODSUMOWANIE

Z przeprowadzonego badania wynika, że zmiany w strukturze pracujących według trzech sektorów w Polsce zmierzają we właściwym kierunku (zgodnie z teorią trzech sektorów), czyli wyraźnie maleje odsetek pracujących w rolnictwie, a rośnie odsetek pracujących w usługach.

W ujęciu regionalnym zmiany w strukturze pracujących według sektorów ekonomicznych w Polsce w latach 2010–2016 były nieznaczne, a udział pracujących w poszczególnych sektorach był zbliżony w większości województw. Dopiero porównanie podobieństwa struktury województw ze strukturą w Polsce oraz rok do roku, a także stopnia zróżnicowania udziału pracujących w sektorach w województwach w pierwszym i ostatnim badanym okresie pozwoliło wskazać te województwa i sektory, w których w ciągu siedmiu lat nastąpiły największe zmiany.

W przypadku podobieństwa sektorowej struktury pracujących w województwie do struktury w Polsce największe podobieństwo zaobserwowano w woj. kujawsko-pomorskim, a najmniejsze – w lubelskim (wyraźna dominacja pracujących w rolnictwie). W największym stopniu w stosunku do struktury Polski zmniejszyło się podobieństwo struktury pracujących w woj. opolskim, a zwiększyło – w małopolskim.

Największe zróżnicowanie udziału pracujących według sektorów w badanym okresie zauważono w woj. lubelskim, a najmniejsze – w małopolskim.

Porównując zmiany w strukturze pracujących według sektorów w ujęciu łańcuchowym, największe zmiany zaobserwowano na początku badanego okresu, czyli w 2011 r. w stosunku do 2010 r. (np. w woj. pomorskim). Natomiast na koniec badanego okresu (porównanie 2016 r. do 2015 r.) w większości województw zaobserwowano znacznie wolniejsze zmiany.

Reasumując, można stwierdzić, że zmiany w strukturze pracujących według sektorów ekonomicznych w województwach zachodzą powoli i charakteryzują się tym, że systematycznie maleje w nich udział pracujących w rolnictwie (sektor 1), a nieznacznie wzrasta udział w sektorze 5, czyli w usługach związanych z działalnością finansową i ubezpieczeniową oraz z obsługą rynku nieruchomości.

BIBLIOGRAFIA

Adamczyk, P. (2012). Regionalne zróżnicowanie przemian w trójsektorowej strukturze osób pracujących w Polsce po akcesji do Unii Europejskiej. *Roczniki Ekonomii Rolnictwa i Rozwoju Obszarów Wiejskich*, *99*(4), 29–37.

Batóg, B., Batóg, J. (2001). Analiza zbieżności struktur zatrudnienia w wybranych krajach wysokorozwiniętych. *Zeszyty Naukowe Uniwersytetu Szczecińskiego*. *Prace Katedry Ekonometrii i Statystyki*, *11*, 155–163.

Dąbrowska, E. (2013). Popyt na pracę w województwie podlaskim w świetle struktury sektorowej gospodarki regionu. *Optimum. Studia Ekonomiczne*, *6*(66), 116–135.

Dworak, E., Malarska, A. (2010). Changes in Sectoral Structure of Employment in Poland in Comparison to the European Union. *Acta Universitatis Lodziensis, Folia Oeconomica*, *242*, 7–29.

Klembowska, D. (2008). Zmiany w zatrudnieniu w Polsce i krajach Unii Europejskiej. *Ekonomika i Organizacja Gospodarki Żywnościowej. Zeszyty Naukowe SGGW*, *72*, 51–63.

Klimczyk, P. (2008). Analiza struktury zatrudnienia w gospodarce polskiej na tle innych państw Unii Europejskiej. *Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie*, *793*, 17–28.

Kosmalski, R. (2008). Teoria trzech sektorów a struktura pracujących w Polsce na tle integracji z Unią Europejską. *Zeszyty Studiów Doktoranckich, Akademia Ekonomiczna w Poznaniu, Wydział Ekonomii*, *44*, 5–24.

Kwiatkowska, W. (2016). Zmiany sektorowej struktury pracujących w Polsce na tle innych krajów Unii Europejskiej w latach 2005–2014. *Studia Prawno-Ekonomiczne*, *99*, 309–326.

Kwiatkowski, E. (1980). *Teoria trzech sektorów gospodarki*. Warszawa: PWN.

Malina, A. (2006). Analiza zmian struktury zatrudnienia w Polsce w porównaniu z krajami Unii Europejskiej. *Prace z zakresu prognozowania. Zeszyty Naukowe Akademii Ekonomicznej w Krakowie*, *726*, 5–21.

Malina, A. (red.) (2008). *Przestrzenno-czasowa analiza rynku pracy w Polsce i krajach Unii Europejskiej*. Kraków: Wydawnictwo Uniwersytetu Ekonomicznego.

Suchecki, B. (2010). *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*. Warszawa: Wydawnictwo C.H. Beck.

Walkowiak, R. (2016). Zmiany struktur zatrudnienia w województwie warmińsko-mazurskim. *Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach*, *258*, 245–252.

Wędrowska, E. (2012). *Miary entropii i dywergencji w analizie struktur*. Olsztyn: Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego.

Zajdel, M. (2010). Trójsektorowa struktura zatrudnienia w polskiej gospodarce jako miernik rozwoju (wybrane aspekty). *Nierówności Społeczne a Wzrost Gospodarczy*, *18*, 419–430.

Wiadomości Statystyczne. The Polish Statistician, 2019, vol. 64, 9, 37–59 DOI: 10.5604/01.3001.0013.7591
Studia interdyscyplinarne. Wyzwania badawcze / Interdisciplinary studies. Research challenges

Dekompozycje czynnikowe przyrostu wartości dodanej brutto według sekcji PKD i województw

Dariusz Kotlewski[a]

Streszczenie. Celem artykułu jest wykazanie, że wykonanie dekompozycji przyrostu wartości dodanej brutto w czterech wariantach umożliwia pogłębienie obserwacji procesów zachodzących w gospodarce. Warianty te uzyskano na podstawie dwóch zasadniczych dychotomii. Pierwsza dotyczyła dekompozycji na kontrybucje wynagrodzeń czynników produkcji z jednej strony oraz na kontrybucje zasobów czynników produkcji i ich produktywności (total factor productivity – TFP) z drugiej strony. Druga polegała na wykonaniu dekompozycji równolegle dla zatrudnionych oraz dla pracujących. Opracowana metodologia umożliwiła dokonanie obliczeń dla lat 2001–2015 na poziomie zagregowanym według sekcji PKD, województw oraz jednocześnie sekcji i województw. Przeprowadzono je na podstawie danych z Banku Danych Lokalnych i rachunków narodowych GUS. Wyniki potwierdzają, że wykonanie dekompozycji według ww. dychotomii umożliwia pogłębienie analiz wzrostu gospodarczego, co jest szczególnie istotne w aspekcie regionalnym.

Słowa kluczowe: czynniki produkcji, wynagrodzenie czynników produkcji, zasoby czynników produkcji, total factor productivity, dekompozycje wartości dodanej brutto, zatrudnieni, pracujący

Factor decompositions of gross value added growth by NACE sections and voivodships

Summary: The aim of this paper is to demonstrate that performing gross value added growth decompositions in four variants makes it possible to deepen the observation of economic processes. These variants have been obtained on the basis of two fundamental dichotomies. The first of them involved performing a decomposition of the gross value added growth into the contributions of production factor remunerations, and, in parallel, a decomposition into the contributions of production factor stocks and total factor productivity (TFP). The second dichotomy involved performing separate but parallel decompositions for employees and for employed persons. The devised methodology made it possible to perform computations for the years 2001–2015 at the aggregate level, according to NACE sections, according to voivodships and according to both the NACE sections and the voivodships. The decompositions were performed basing on Statistics Poland's data from the Bank of Local Data and the National Accounts. The presented results confirm that performing decompositions according to the two above-mentioned dichotomies makes it possible to deepen the analyses of the economic growth, which is especially important in the regional aspect.

Keywords: production factors, factor remunerations, factor stocks, total factor productivity, gross value added growth decompositions, employees, employed persons

JEL: O47, E22, E23, E24

[a] Szkoła Główna Handlowa, Kolegium Nauk o Przedsiębiorstwie.

Celem artykułu jest wykazanie, że wykonanie dekompozycji w czterech zaprezentowanych wariantach umożliwia pogłębienie obserwacji procesów gospodarczych. Pozwala to na przeprowadzanie analiz idących dalej niż dotychczasowe.

Niniejszy artykuł dotyczy dekompozycji stanowiących istotne rozwinięcie w stosunku do projektu zaprezentowanego w pracy Kotlewskiego (2017a), który dotyczył dekompozycji przyrostu wartości dodanej brutto (WDB) na kontrybucje wynagrodzeń czynników produkcji dla województw i sektorów gospodarki polskiej (sekcji PKD). Rozbudowano techniki rachunkowe związane z przygotowaniem danych do postaci nadającej się do wykorzystania w rachunku dekompozycji w taki sposób, że możliwe stało się wykonanie dekompozycji przyrostu wartości dodanej brutto na kontrybucje zasobów czynników pracy i kapitału oraz total factor productivity (TFP). Ta ostatnia postać rachunku dekompozycji bardziej odpowiada neoklasycznej teorii wzrostu gospodarczego Solowa (1956) i tzw. dekompozycji Solowa (1957)[1]. Kolejnym rozwinięciem jest wykonanie rachunków w odniesieniu do zatrudnionych oraz pracujących (według definicji tych kategorii obowiązujących w statystykach GUS), co też jest następstwem doskonalenia technik związanych z przygotowaniem danych. Przeprowadzenie dekompozycji równolegle dla zatrudnionych i pracujących umożliwia z kolei wykonywanie dodatkowych porównań. Poprzednia edycja projektu zaprezentowanego w pracy Kotlewskiego (2017a) została opracowana jako wkład własny autora pt. *Dekompozycje czynnikowe WDB na zatrudnionego* do pracy Lewandowskiego (2015)[2]. Podobnie obecna edycja projektu została udostępniona jako wkład własny autora pt. *Dekompozycja czynnikowa wartości dodanej brutto*[3].

CHARAKTERYSTYKA OGÓLNA

Punktem wyjścia jest stwierdzenie (Kotlewski, 2017a), że dekompozycja czynnikowa polega na rozdzieleniu wzrostu gospodarczego, w ujęciu względnym, na kontrybucje czynników produkcji (zwykle pracy i kapitału), określanych niekiedy jako pierwotne[4], przy czym w rachunkowości wzrostu gospodarczego

[1] Zob. dalej równanie (2) i związany z nim komentarz. Ideą dekompozycji Solowa jest wydzielenie wkładów do wzrostu gospodarczego (czyli kontrybucji) czynników produkcji, tj. pracy i kapitału. Nieoczekiwanie Robert Solow odkrył, że oprócz wydzielonych wkładów czynników produkcji pojawił się dodatkowy wkład jakby nieznanego czynnika, który został przez niego zinterpretowany jako postęp techniczny.

[2] Badanie przeprowadzone w ramach projektu unijnego POPT I (Program Operacyjny Pomoc Techniczna – edycja pierwsza), zob. Kotlewski (2015).

[3] Badanie przeprowadzone w ramach projektu unijnego POPT II (Program Operacyjny Pomoc Techniczna – edycja druga), zob. Kotlewski (2018a i 2018b).

[4] Na przykład przez Hultena (2009).

obecnie preferuje się WDB jako miarę poziomu aktywności gospodarczej, m.in. z uwagi na stosowanie w rachunkach narodowych równości[5]:

$$WDB = WP + WK \tag{1}$$

gdzie:
WP – całkowite wynagrodzenie pracy (na poziomie danej agregacji),
WK – wynagrodzenie kapitału.

Przyjmuje się zgodnie z teorią ekonomii, że czynniki produkcji są wynagradzane według ich krańcowych produktywności, dlatego analiza ich wynagrodzeń zasadniczo stanowi również, w świetle tego założenia, analizę ich produktywności. Takie analizy przeprowadzono w pierwszej edycji projektu (POPT I).

Badanie kontrybucji[6] wynagrodzeń czynników zrealizowane w edycji POPT I nie zawiera jednak wydzielenia jakości tych czynników i ich odrębnej analizy. Na przykład kontrybucja pracy rozumiana jako kontrybucja wynagrodzenia pracy (*WP*) obejmuje nie tylko kontrybucję samego nakładu pracy, mierzonej jako liczba zatrudnionych, liczba pracujących, liczba etatów ekwiwalentnych albo liczba godzin przepracowanych, czyli w postaci jednej z dopuszczalnych przez teorię definicji zasobu czynnika pracy (OECD, 2001, s. 40–41), lecz także kontrybucję wydajności pracy (*labour efficiency*) oraz kontrybucję wykorzystania zasobu pracy (*labour capacity utilisation*) (Havik i in., 2014, s. 10 i 32). Podobnie kontrybucja kapitału, rozumiana jako kontrybucja wynagrodzenia kapitału (*WK*), obejmuje nie tylko kontrybucję samego nakładu kapitału, mierzonego jako jego zasób w postaci np. stanu środków trwałych[7], tj. jego kolejnych przybliżeń w różnych cenach, ewentualnie uzupełniony o wartości niematerialne i prawne, itd., lecz także kontrybucję wydajności kapitału (*capital efficiency*) oraz kontrybucję wykorzystania zasobu kapitału (*capital capacity utilisation*) (Havik i in., 2014, s. 10 i 32). Wydajność czynników produkcji identyfikuje się zwykle jako mającą związek z długookresowym postępem organizacyjnym i technicznym. Z kolei wykorzystanie zasobów czynników produkcji identyfikuje się zwykle jako mające

[5] Rachunki narodowe oparte są na systemach SNA (System of National Accounts) oraz ESA (European System of Accounts). Równanie (1) jest spełnione, jeżeli się założy, że w gospodarce panuje doskonała konkurencja, a przychody skali są stałe. PKB różni się od WDB tym, że obejmuje także podatki od produktów, ale nie obejmuje subsydiów do produktów. Ze względów rachunkowych (ścisłość metodologiczna) w dekompozycjach wzrostu gospodarczego preferuje się WDB, a nie PKB.

[6] W rachunkowości wzrostu gospodarczego przyjęło się już określenie *kontrybucja*, rozumiane jako wkład do wzrostu gospodarczego, którego najodpowiedniejszym reprezentantem (ze względów metodologicznych, o czym dalej) jest przyrost WDB.

[7] Przy pewnych założeniach dotyczących sposobu liczenia zasobu kapitału (chodzi o metodę ciągłej inwentaryzacji według założeń przyjętych np. w rachunku KLEMS, mających zastosowanie również tutaj) jego przyrost jest równy przyrostowi usług kapitału.

związek głównie z cyklem koniunkturalnym. Zarówno wydajność czynników, jak i wykorzystanie ich zasobów traktuje się jako komponenty tzw. reszty Solowa, czyli TFP, tj. produktywności łącznej czynników.

Wyznaczenie TFP niesie ze sobą zatem pewne dodatkowe zasoby informacji dla analizy wzrostu gospodarczego i dlatego, oprócz wzoru wyjściowego (1), w edycji POPT II wykorzystano także dekompozycję Solowa (1957) o wzorze ogólnym:

$$\Delta Y/Y = \alpha\,\Delta L/L + (1-\alpha)\,\Delta K/K + \Delta A/A \tag{2}$$

gdzie:
Y – poziom aktywności gospodarczej (zwykle PKB lub WDB),
L – zasób czynnika pracy,
K – zasób czynnika kapitału,
A – produktywność łączna czynników, czyli TFP.

Udziały czynników w gospodarce α oraz $\beta = 1 - \alpha$ obliczane są jako udziały wynagrodzeń tych czynników w PKB lub WDB.

Aby równanie (1) było zawsze formalnie spełnione, jedna z trzech użytych w niej zmiennych musi być obliczana rezydualnie, czyli jako odpowiednia różnica pomiędzy pozostałymi wartościami. Najczęściej w rachunkowości wzrostu gospodarczego (np. w dekompozycjach rachunku produktywności KLEMS[8]) oblicza się rezydualnie *WK*, ponieważ dane dotyczące tej zmiennej są trudniejsze do zmierzenia lub oszacowania niż dane dotyczące odpowiedniej zmiennej związanej z *WP*. Stąd w edycji POPT I zastosowano metodę rezydualnie obliczanych zmiennych (zarówno poziomów, przyrostów, jak i kontrybucji, tak aby równania były dokładnie spełnione w każdej postaci) dotyczących wynagrodzenia kapitału. Uznano tę metodę za lepszą od ewentualnie możliwej metody rezydualnie obliczanych zmiennych dotyczących wynagrodzenia pracy. Wyniki obydwu powinny być zbliżone, przynajmniej na poziomie wyższych agregacji. Jednak w równaniu (2) zmienna kapitałowa *K* (pomimo, że udział $\beta = 1 - \alpha$ jest również obliczany rezydualnie) musi być zasilona egzogenicznie do rachunku (czyli przez pobranie do rachunku dodatkowych danych statystycznych, chociaż z możliwymi ich przekształceniami), oprócz zmiennej *L*, aby możliwe było obliczenie w sposób rezydualny wartości związanej z symbolem *A*, który reprezentuje TFP. Wykonanie takiego rachunku wymagało odpowiedniego przygotowania danych dotyczących tej dodatkowej zmiennej, co wiązało się z dalszą pracą

[8] Rachunek ten był rozwijany przez Jorgensona (1963, 1989), Jorgensona, Gollopa i Fraumeni (1987), Jorgensona i Grilichesa (1967) oraz Jorgensona, Ho i Stiroha (2005). Podsumowanie metodologii w wersji europejskiej EU KLEMS można znaleźć w: O'Mahony i Timmer (2009) oraz Timmer i in. (2007).

metodologiczną na poziomie przetwarzania danych wejściowych do rachunku. Dane dotyczące zmiennej *K* nie są dostępne jednocześnie w podziale na sekcje PKD i województwa w zasobach GUS. Z tego powodu w obecnej edycji projektu POPT zrealizowano również dekompozycję bazującą na wzorze ogólnym (2). Wykonano również pracę metodologiczną mającą na celu konwersję czynnika pracy z zatrudnionych na pracujących, co pozwoliło na dwutorową realizację rachunku dla obydwu wyróżnionych kategorii, czyli według dwóch z czterech ww. dopuszczalnych definicji czynnika pracy[9].

W porównaniu z rachunkiem produktywności KLEMS przeprowadzonym także w GUS (Kotlewski i Błażej, 2016, 2018) niniejsze rachunki dekompozycji nie uwzględniają jakości pracy jako dodatkowego czynnika produkcji ani podziału na niższe agregacje sektorowe niż sekcje PKD[10]. Dzięki dostępności danych GUS stało się jednak możliwe wykonanie dekompozycji według województw. Wszystkie dane wejściowe do rachunków są w klasyfikacji PKD 2007, o ile zostały przeliczone w GUS. W niektórych przypadkach wykorzystano dane z PKD 2004, w sytuacji gdy ich przeliczenie na PKD 2007 nie zostało jednak w GUS zrealizowane[11].

OPIS SFORMALIZOWANY

Rachunki na agregatach makroekonomicznych

W edycji POPT I omawianego badania (która w edycji POPT II stała się częścią większej całości) zamiast równania (1) wykorzystano do tego celu następujące równanie:

$$\Delta WDB/WDB_{(-1)} = \alpha \Delta WP/WP_{(-1)} + \beta \Delta WK/WK_{(-1)} \qquad (3)$$

[9] Według OECD (2001, s. 40 i 41) najlepszą miarę stanowią godziny przepracowane, ale dane dotyczące godzin przepracowanych są dostępne w GUS tylko dla całej gospodarki, a nie w ujęciu województw. Obserwacja danych wynikowych w trakcie prac nad rachunkiem wykazała, że różnice pomiędzy wynikami uzyskanymi dla osób zatrudnionych lub pracujących z jednej strony, a dla godzin przepracowanych przez zatrudnionych lub pracujących – z drugiej, są zaniedbywalne małe i nieistotne dla analiz (wszędzie tam, gdzie jest możliwe uzyskanie wyników według obu wersji, czyli przede wszystkim dla całej gospodarki narodowej, jak w rachunku KLEMS).

[10] Ponadto w rachunku produktywności KLEMS jest mowa raczej o usługach czynników produkcji, a nie ich zasobach, co wynika ze stosowania w tym rachunku procedury Törnqvista. Zagadnienie to wykracza jednak poza zakres niniejszych rozważań metodologicznych i nie ma istotnego wpływu na niniejsze analizy.

[11] Przeliczenia takie wymagają bardzo dużych nakładów pracy różnych zespołów w GUS i w niezbędnej części zostały już dokonane na potrzeby rachunków narodowych. Dane, które pozostały tylko w PKD 2004, raczej nie będą już przeliczane. Konieczność ich wykorzystania w niektórych sytuacjach nie ma jednak dużego znaczenia jakościowego dla wyników z punktu widzenia ich analizy. Dotyczy to m.in. wielu danych w Banku Danych Lokalnych (BDL).

gdzie:
$\alpha = (WP/WDB + WP_{(-1)}/WDB_{(-1)})/2$,
$\beta = (WK/WDB + WK_{(-1)}/WDB_{(-1)})/2$,
indeks (–1) – wartości za rok poprzedni.

Jeżeli według rachunków narodowych obowiązuje równanie (1), to przyrost względny (procentowy) wartości dodanej brutto $\Delta WDB/WDB_{(-1)}=(WDB-WDB_{(-1)})/WDB_{(-1)}$ jest równy sumie przyrostów względnych (procentowych) wynagrodzenia pracy $\Delta WP/WP_{(-1)}=(WP-WP_{(-1)})/WP_{(-1)}$ oraz wynagrodzenia kapitału $\Delta WK/WK_{(-1)}=(WK-WK_{(-1)})/WK_{(-1)}$, zważonych udziałami tych czynników w WDB. Zależności te są spełnione, jeżeli opisane przyrosty są infinitezymalnie małe, czyli w czasie ciągłym. Jeżeli czas nie jest traktowany jako ciągły, tylko jako dyskretny, tzn. gdy występują mierzalne interwały czasowe, wówczas należy stosować wagi α i β w postaci wzorów – podanych w komentarzu do wzoru (3) – na średnie międzyokresowe udziały czynników produkcji w WDB. Dokonuje się więc interpolacji liniowej udziałów pomiędzy okresami bieżącym i poprzednim. Oznacza to, że z powodu stosowania czasu dyskretnego dwa wyrażenia po prawej stronie wzoru (3) są obarczone pewnym niewielkim odchyleniem od nieznanych wartości rzeczywistych, a więc jest to równanie przybliżone, ponieważ interpolacja liniowa jest procedurą przybliżoną. Aby to odchylenie nie narastało przy dalszych obliczeniach, przyjmuje się dla kontrybucji wynagrodzenia kapitału (KWK), zamiast $\beta\Delta WK/WK_{(-1)}$, wartość rezydualną według wzoru:

$$KWK = \Delta WDB/WDB_{(-1)} - \alpha\Delta WP/WP_{(-1)} \tag{4}$$

czyli kontrybucję wynagrodzenia kapitału w przyroście WDB oblicza się poprzez odjęcie od przyrostu tej ostatniej kontrybucji wynagrodzenia pracy. Podobnie postępuje się dalej z innymi kontrybucjami kapitału, co zapewnia ścisłość formalną i bilansowanie się rachunku. W edycji POPT II przeprowadzono te obliczenia dwutorowo, równolegle dla liczby zatrudnionych Z i liczby pracujących P, tj. dla WP oznaczanego jako WP_Z albo WP_P i odpowiednio WK oznaczanego jako WK_Z albo WK_P (gdyż te wartości w pewnym stopniu się różnią także dla tych zmiennych kapitałowych). Oznacza to również, że udział α jest w pewnym stopniu inny dla zatrudnionych i pracujących i w związku z tym jest oznaczany jako α_z albo α_p z czego wynika, że także β należy oznaczać jako β_z albo β_p. Wszystkie udziały (wagi) zostały odrębnie wyznaczone dla każdej agregacji, tj. dla całej gospodarki polskiej, dla województw, dla sekcji oraz jednocześnie dla sekcji i województw, z wyjątkiem wag dla niektórych odchyleń od średniej krajowej – zob. s. 45. Dla przejrzystości w tym miejscu pominięto we wzorach indeksy z tym związane.

W rachunku z uwzględnieniem wyznaczenia TFP w edycji POPT II przekształcono wzór (3) do postaci inspirowanej przez wzór ogólny (2):

$$\Delta WDB/WDB_{(-1)} = \frac{\alpha_z \Delta Z}{Z_{(-1)}} + \frac{\beta_z \Delta K}{K_{(-1)}} + \frac{\Delta TFP_z}{TFP_{z(-1)}} \tag{5}$$

albo:

$$\Delta WDB/WDB_{(-1)} = \frac{\alpha_p \Delta P}{P_{(-1)}} + \frac{\beta_p \Delta K}{K_{(-1)}} + \frac{\Delta TFP_P}{TFP_{P(-1)}} \tag{6}$$

wstawiając jednocześnie odpowiednie wartości *Z* albo *P* dla zatrudnionych i pracujących oraz indeksy z nimi związane[12]. Jak wynika ze wzorów (5) i (6), zmienna kapitałowa *K* nie zmienia się w zależności od czynnika pracy, dlatego nie jest indeksowana. Natomiast udział kapitału w wartości dodanej β zmienia się, przyjmując wartości odpowiednio β_z albo β_p. W przeciwieństwie do wzoru (3), dla którego przyjęcie sposobu obliczania podanego we wzorze (4) zwalnia z konieczności wyznaczenia β, tutaj konieczne jest przyjęcie założenia, że $\beta = 1-\alpha$ (z odpowiednimi indeksami *Z* albo *P*), aby równania (5) i (6) były rozwiązywalne. Udziały czynników we wzorach (5) i (6) są takie same jak we wzorze (3), który również w kompletnym zapisie powinien uwzględniać indeksy *Z* albo *P*. Wszystkie pozostałe zmienne po prawej stronie równań muszą być zawsze obliczane albo tylko dla zatrudnionych *Z*, albo tylko dla pracujących *P*[13]. Zmienna związana z TFP, reprezentująca według teorii wkład postępu technicznego i organizacyjnego we wzrost gospodarczy, czyli kontrybucja TFP (*KTFP*), jest obliczana rezydualnie, podobnie jak zmienna kapitałowa w równaniu (4), czyli według wzoru dla zatrudnionych:

$$KTFP = \frac{\Delta WDB}{WDB_{(-1)}} - \alpha_z \frac{\Delta Z}{Z_{(-1)}} - \beta_z \frac{\Delta K}{K_{(-1)}} \tag{7}$$

Dla pracujących we wzorze (7) należy wstawić odpowiednie wartości i indeksy dolne *P* zamiast *Z.*

Aby równanie reprezentowane przez wzór (7) było rozwiązywalne, tutaj także konieczne było przyjęcie podstawowego założenia neoklasycznego o stałych przychodach skali w warunkach doskonałej konkurencji, czyli że $\beta = 1-\alpha$, oczywiście z odpowiednimi indeksami *Z* albo *P*. Założono, że zasada ta będzie obowiązywać wszędzie dalej.

[12] W tym miejscu podano wzory dla zatrudnionych i pracujących dla przykładu, dalej – tylko dla jednej z tych wersji.

[13] Faktycznie jest $TFP_Z \approx TFP_P$, ale przybliżenie to jest stosunkowo niedokładne ze względów narzędziowych, dlatego zachowuje się indeksy *Z* i *P*.

Rachunki dla wynagrodzeń czynników produkcji na osobę biorącą udział w procesie produkcyjnym

Dla przyrostów na zatrudnionego równanie teoretyczne (3) należy przekształcić do postaci:

$$\frac{\Delta(WDB/Z)}{WDB_{(-1)}/Z_{(-1)}} = \alpha_z \frac{\Delta(WP_Z/Z)}{WP_{Z(-1)}/Z_{(-1)}} + \beta_z \frac{\Delta(WK_Z/Z)}{WK_{Z(-1)}/Z_{(-1)}} \tag{8}$$

gdzie:
Z – liczba zatrudnionych w okresie bieżącym,
$Z_{(-1)}$ – liczba zatrudnionych w okresie poprzednim.

Odpowiednio w równaniu dotyczącym pracujących stosuje się wartości i indeks P[14]. W równaniu (8) obowiązują zależności $\Delta(WDB/Z) = WDB/Z - WDB_{(-1)}/Z_{(-1)}$, $\Delta(WP_Z/Z) = WP_Z/Z - WP_{Z(-1)}/Z_{(-1)}$ i $\Delta(WK_Z/Z) = WK_Z/Z - WK_{Z(-1)}/Z_{(-1)}$. Jednak w praktyce kontrybucję wynagrodzenia kapitału do przyrostu WDB na zatrudnionego (KWK_Z) wyznacza się rezydualnie – zamiast stosować wyrażenie $\beta_z\Delta(WK_Z/Z)/(WK_{Z(-1)}/Z_{(-1)})$ – zgodnie z równaniem (odpowiednio dla pracujących symbol Z należy zamienić na P):

$$KWK_Z = \frac{\Delta(WDB/Z)}{WDB_{(-1)}/Z_{(-1)}} - \alpha_z \frac{\Delta(WP_Z/Z)}{WP_{Z(-1)}/Z_{(-1)}} \tag{9}$$

Z kolei odchylenia WDB na zatrudnionego dla danego województwa, sekcji lub województwa i sekcji jednocześnie (albo ewentualnie innych wybranych agregacji), w stosunku do średniej krajowej oraz kontrybucje czynników do tego odchylenia teoretycznie spełniają równanie:

$$\frac{WDB_j/Z_j - WDB/Z}{WDB/Z} = \alpha_z \frac{WP_{Z,j}/Z_j - WP_Z/Z}{WP_Z/Z} + \beta_z \frac{WK_{Z,j}/Z_j - WK_Z/Z}{WK_Z/Z} \tag{10}$$

gdzie indeks j odnosi się do wartości dla danego województwa, sekcji lub województwa i sekcji jednocześnie (lub ewentualnie innych wybranych agregacji), podczas gdy pozostałe wartości są wartościami dla całego kraju (odpowiednie wartości i indeksy P należy zastosować dla pracujących). Tak jak w przypadku poprzednich wzorów kontrybucję wynagrodzenia kapitału do odchylenia WDB na zatrudnionego ($KWKO_Z$) oblicza się rezydualnie ze wzoru:

[14] Uwaga dotyczy także kolejnych równań.

$$KWKO_Z = \frac{WDB_j/Z_j - WDB/Z}{WDB/Z} - \alpha_z \frac{WP_{Z,j}/Z_j - WP_Z/Z}{WP_Z/Z} \tag{11}$$

zamiast stosować wyrażenie $\beta_Z(WK_{Z,j}/Z_j - WK_Z/Z)/(WK_Z/Z)$.

Wagi α_z i β_z są tu obliczane z następujących wzorów $\alpha_z = WP_Z/WDB$ oraz $\beta_z = WK_Z/WDB$ (podobnie dla pracujących). Nie oblicza się więc ich w drodze interpolacji liniowej pomiędzy udziałami z dwóch okresów, jak w pozostałych przypadkach, ponieważ zawsze używa się danych we wzorach (10) i (11) tylko z jednego okresu[15]. Ze względu na czytelność wzorów zrezygnowano tutaj z dodawania kolejnych indeksów.

Generalnie w całym rachunku zastosowano średnie arytmetyczne wag z dwóch okresów, ubiegłego i bieżącego (na każdym poziomie agregacji), co jest wzorowane na procedurze Törnqvista, związanej z porównywaniem dwóch okresów lub dwóch sytuacji. Jednak dla odchyleń wszystkie dane pochodzą z jednego, a nie z dwóch okresów, natomiast porównuje się dwie jednostki w tym samym czasie, czyli dwie sytuacje. Gdyby zatem porównywano dwa województwa, należałoby zastosować średnie arytmetyczne wagi dla obu porównywanych województw. Tutaj jednak porównuje się dwie jednostki z różnych poziomów taksonomicznych, tj. województwa do całej Polski. Dla odchyleń zdecydowano się zastosować wagi dla całego kraju, traktując wagi dla Polski jako właściwy punkt odniesienia, przy zachowaniu zróżnicowania wag według sekcji PKD.

Jako uzupełnienie powyższych rachunków wykonano dekompozycję zmian w odchyleniu od średniej krajowej, pozwalającą wyraźniej zaobserwować, czy różnica w stosunku do średniej krajowej powiększa się czy zmniejsza. Rachunki te wymagają dodania w odpowiednich miejscach symbolu Δ we wzorach (10) i (11).

Rachunki na osobę biorącą udział w procesie produkcyjnym z wyznaczeniem TFP

Przy przeprowadzaniu powyższych działań w sposób inspirowany równaniem (2), w celu obliczenia wartości związanych z TFP, równanie (8) należy zastąpić równaniem dla zatrudnionych Z:

$$\frac{\Delta(WDB/Z)}{WDB_{(-1)}/Z_{(-1)}} = \alpha_z \frac{\Delta(Z/Z)}{Z_{(-1)}/Z_{(-1)}} + \beta_z \frac{\Delta(K/Z)}{K_{(-1)}/Z_{(-1)}} + \frac{\Delta(TFP_Z/Z)}{TFP_{Z(-1)}/Z_{(-1)}} \tag{12}$$

[15] To skomplikowane zagadnienie zostało tu potraktowane w sposób uproszczony (zob. Milana, 2009).

albo odpowiednim równaniem dla pracujących *P*, w którym wartości i indeks *Z* należy zastąpić wartościami i indeksem *P*[16].

W tym wypadku zachodzi pewna osobliwość, która polega na tym, że wartości związane z czynnikiem pracy ulegają skróceniu. Stąd równanie (12) upraszcza się do postaci:

$$\frac{\Delta(WDB/Z)}{WDB_{(-1)}/Z_{(-1)}} = \beta_z \frac{\Delta(K/Z)}{K_{(-1)}/Z_{(-1)}} + \frac{\Delta(TFP_Z/Z)}{TFP_{Z(-1)}/Z_{(-1)}} \tag{13}$$

Przyjęcie neoklasycznego założenia o stałych przychodach skali w warunkach doskonałej konkurencji pozwala tutaj zastosować wzór $\beta = 1-\alpha$, z odpowiednimi indeksami *Z* i *P,* co czyni równanie (13) rozwiązywalnym. To z kolei umożliwia wyznaczenie kontrybucji TFP ($KTFP_Z$), tym razem z konieczności[17], w sposób rezydualny:

$$KTFP_Z = \frac{\Delta(WDB/Z)}{WDB_{(-1)}/Z_{(-1)}} - (1-\alpha_z)\frac{\Delta(K/Z)}{K_{(-1)}/Z_{(-1)}} \tag{14}$$

Z kolei dla odchyleń od średniej krajowej wzór (10) należy w tym wypadku zastąpić wzorem:

$$\frac{WDB_j/Z_j - WDB/Z}{WDB/Z} = \alpha_z \frac{Z_j/Z_j - Z/Z}{Z/Z} + \beta_z \frac{K_j/Z_j - K/Z}{K/Z} + \frac{TFP_{Z,j}/Z_j - TFP_Z/Z}{TFP_Z/Z} \tag{15}$$

W przypadku tego równania także zachodzi osobliwość związana z czynnikiem pracy, pozwalająca uprościć je do postaci:

$$\frac{WDB_j/Z_j - WDB/Z}{WDB/Z} = \beta_z \frac{K_j/Z_j - K/Z}{K/Z} + \frac{TFP_{Z,j}/Z_j - TFP_Z/Z}{TFP_Z/Z} \tag{16}$$

Przyjęcie neoklasycznego założenia o stałych przychodach skali w warunkach doskonałej konkurencji tutaj pozwala zastosować wzór $\beta = 1-\alpha$, z odpowiednimi indeksami *Z* i *P*, co z kolei umożliwia wyznaczenie kontrybucji TFP ($KTFPO_Z$) w sposób rezydualny:

$$KTFPO_Z = \frac{WDB_j/Z_j - WDB/Z}{WDB/Z} - (1-\alpha_z)\frac{K_j/Z_j - K/Z}{K/Z} \tag{17}$$

[16] Uwaga dotyczy również kolejnych równań.

[17] W przeciwieństwie do opcjonalnego rezydualnego wyznaczania zmiennych związanych z kapitałem we wzorach: (1), (3), (4), (8), (9), (10) i (11).

Jako uzupełnienie powyższych rachunków wykonano dekompozycję zmian w odchyleniu od średniej krajowej, co wymagało dodania w odpowiednich miejscach symbolu *Δ* we wzorach (16) i (17). W tych rachunkach wagi dla odchyleń inaczej policzono, podobnie jak w przypadku wzorów (10) i (11).

Ze względu na wpływ inflacji na pomiary przyrostów w czasie w powyższych wzorach należy użyć wartości realnych, czyli od wartości bieżącej w cenach stałych (z ubiegłego roku) odejmujemy wartość z ubiegłego roku w cenach bieżących z tego roku.

WYZWANIA ZWIĄZANE Z PRZYGOTOWANIEM DANYCH

W zakresie potrzebnym do pierwszej edycji projektu (POPT I) sposób przetworzenia danych został podany w pracy Kotlewskiego (2017b) wraz z wyjaśnieniem przesłanek koncepcyjnych takiego postępowania. W drugiej edycji projektu metody te znacznie rozbudowano. Tutaj są one przedstawione łącznie.

Do obliczenia rezydualnej rentowności kapitału brutto według wzoru *RK* = *WK/K* z odpowiednimi indeksami niezbędne są dane dotyczące stanu środków trwałych netto w cenach bieżących[18]. Tymczasem dane BDL GUS są wyrażone w cenach ewidencyjnych, czyli w cenach nominalnych z wielu okresów poniesienia wydatków inwestycyjnych na środki trwałe, przy czym są to środki trwałe brutto, czyli bez uwzględnienia deprecjacji kapitału (zwykle utożsamianej z amortyzacją). W zasobach GUS dostępne są także dane dotyczące stanu środków trwałych netto w cenach bieżących według sekcji, ale bez podziału na województwa[19], dlatego dane z BDL posłużyły jedynie jako struktura[20] do oszacowania danych w cenach bieżących dla województw według wzoru:

$$KN_{bsw} = \frac{KB_{esw}}{KB_{es}} KN_{bs} \tag{18}$$

gdzie:

KN_{bsw} – obliczone środki trwałe netto (kapitał netto) w cenach bieżących według sekcji PKD i województw,

KB_{esw} – środki trwałe brutto (kapitał brutto) w cenach ewidencyjnych według sekcji PKD i województw (dane z BDL),

[18] W ramach POPT I wykonano oprócz dekompozycji pewne rachunki wstępne dotyczące udziału wynagrodzenia pracy w WDB potrzebne do wyznaczenia *α* oraz tempa zmian tego udziału, a także obliczono rezydualną rentowność kapitału brutto wraz z jej tempem.

[19] Dane z tablic transmisyjnych (TT) do Eurostatu.

[20] Struktura ta ma postać wektora lub macierzy tablicowej (w tym wypadku o wymiarach sekcje *s* na województwa *w*) odrębnej dla każdego roku, w której wpisane są odpowiednie relacje (w tym wypadku KB_{esw}/KB_{es}) inne dla każdej sekcji i każdego województwa jednocześnie (czyli tych relacji jest *s* razy *w*) i przez którą przemnażane są wektory (w tym wypadku KN_{bs}) lub inne macierze tablicowe odrębne dla każdego roku. Dotyczy to również dalszych wzorów.

KB_{es} – środki trwałe brutto w cenach ewidencyjnych według sekcji PKD dla Polski (dane z BDL),
KN_{bs} – środki trwałe netto w cenach bieżących według sekcji PKD dla Polski (dane z TT).

Choć operacja ta została przeprowadzona jeszcze podczas edycji POPT I, to w edycji POPT II zyskała na znaczeniu, ponieważ w rachunkach dekompozycji czynnikowej z wyznaczeniem TFP konieczne jest wykorzystanie wielkości związanych ze stanem środków trwałych, czyli zasobem kapitału *K*, zamiast wynagrodzenia kapitału *WK* pozyskiwanego z innych źródeł i niewymagającego przeprowadzania tej operacji.

W tym celu, dodatkowo, w drugiej edycji projektu trzeba było uzyskać wartości kapitału netto także w cenach stałych. Przemnożenie KN_{bsw} ze wzoru (18) przez stosunek KN_{ss}/KN_{bs}, gdzie KN_{ss} to środki trwałe netto w cenach stałych według sekcji PKD dla Polski (dane z TT), pozwala otrzymać KN_{ssw}, czyli obliczone środki trwałe netto (kapitał netto) w cenach stałych według sekcji PKD i województw:

$$KN_{ssw} = \frac{KN_{ss}}{KN_{bs}} KN_{bsw} \tag{19}$$

Dla oczyszczenia zmian wartości z efektów inflacyjnych i uzyskania przyrostów realnych potrzebne są także inne zmienne w cenach stałych. Operację tę wykonano dla wartości dodanej brutto, wynagrodzenia pracy oraz wynagrodzenia kapitału. W TT do Eurostatu są podane dane dla tych wartości w podziale na sekcje PKD dla całej gospodarki, w cenach bieżących i cenach stałych. Dane te nie występują jednak w podziale na województwa i dlatego posłużyły tylko jako struktura do wyszacowania WDB w cenach stałych według wzoru:

$$WDB_{ssw} = \frac{WDB_{ss}}{WDB_{bs}} WDB_{bsw} \tag{20}$$

gdzie:
WDB_{ssw} – obliczona wartość dodana brutto w cenach stałych według sekcji PKD i województw,
WDB_{ss} – wartość dodana brutto w cenach stałych według sekcji PKD dla Polski (dane z TT),
WDB_{bs} – wartość dodana brutto w cenach bieżących według sekcji PKD dla Polski (dane z TT),
WDB_{bsw} – wartość dodana brutto w cenach bieżących według sekcji PKD i województw (dane z BDL).

Także ta operacja została zrealizowana w pierwszej edycji projektu, ale w edycji POPT II ma znaczenie podstawowe.

W TT podane są ceny stałe i bieżące dla WDB. Jednakże w przypadku wynagrodzenia pracy dane dostępne są tylko w cenach bieżących. Przyjmuje się, że w odróżnieniu od wartości dodanej, dla której inflacja jest różna w zależności od sekcji PKD (ponieważ poziom produkcji w ujęciu wartościowym dla danej sekcji jest związany z inflacją w danej sekcji), inflacja dla rynku pracy powinna być traktowana jako średnia ważona, ponieważ zatrudnieni są odbiorcami szerokiego koszyka towarów z wielu sekcji[21]. Do oszacowania brakujących wartości zastosowano zatem wzór:

$$WP_{ssw} = \frac{WDB_s}{WDB_b} WP_{bsw} \tag{21}$$

gdzie:

WP_{ssw} – obliczone wynagrodzenie pracy w cenach stałych według sekcji PKD i województw,

WDB_s – wartość dodana brutto w cenach stałych dla Polski (dane z TT),

WDB_b – wartość dodana brutto w cenach bieżących dla Polski (dane z TT),

WP_{bsw} – wynagrodzenie pracy w cenach bieżących według sekcji PKD i województw (dane z BDL).

Użyto przy tym wartości WDB dla Polski, ponieważ wartości dla województw w cenach stałych są już oszacowane proporcjonalnie z wartości dla Polski. Również ta operacja została wykonana w POPT I, ale dla obecnej edycji ma znaczenie podstawowe.

W przypadku realnego wynagrodzenia kapitału zastosowano równanie (1) przekształcone do postaci $WK_{ssw} = WDB_{ssw} – WP_{ssw}$, czyli realne wynagrodzenie kapitału obliczono dla każdej sekcji i dla każdego województwa rezydualnie, jako różnicę pomiędzy realną wartością dodaną brutto a realnym wynagrodzeniem pracy.

Z kolei wszystkie przyrosty policzono jako stosunki różnic pomiędzy wartościami realnymi dla roku bieżącego a wartościami bieżącymi dla roku poprzedniego w odniesieniu do wartości bieżących z roku poprzedniego. Dzięki temu obliczanie wartości w cenach stałych z roku bazowego (np. 2010 r.) stało się zbędne, gdyż otrzymuje się szeregi czasowe dla przyrostów względnych realnych bez tej procedury.

[21] Właściwie inflację dla rynku pracy powinno się liczyć w zależności od grup zatrudnionych według ich zamożności, zamieszkania, wieku i innych podziałów rynku pracy. Obecnie jednak nie jest to możliwe do wykonania z uwagi na brak odpowiednich danych.

Nowym wyzwaniem metodologicznym było wyznaczenie liczby pracujących do rachunku dekompozycji. Dane dostępne w zasobach BDL i spełniające wymagania projektu dotyczą jedynie zatrudnionych. Są one udostępnione w podziale na sekcje i województwa i odpowiadają przyjętym w projekcie grupowaniom klasyfikacyjnym, w tym także dla szeregów czasowych przyjętych w projekcie. Natomiast dane na temat pracujących cechują się mocno skróconymi szeregami czasowymi oraz są dostępne dla grup sekcji (nie zaś dla poszczególnych sekcji) zupełnie innych niż przyjęte w tej części projektu. W związku z powyższym potrzebna okazała się konwersja danych dotyczących zatrudnionych na odpowiadające im dane dotyczące pracujących. Operację tę przeprowadzono według wzoru:

$$P_{sw} = Z_{sw}\frac{P_s}{Z_s} \tag{22}$$

gdzie:
P_{sw} – obliczona liczba pracujących według sekcji PKD i województw,
Z_{sw} – liczba zatrudnionych według sekcji i województw (dane z BDL),
Z_s – liczba zatrudnionych według sekcji dla Polski (dane z BDL),
P_s – liczba pracujących według sekcji dla Polski (dane z TT).

Pozwoliło to określić zasób czynnika pracy (definiowanego jako pracujący) dla wszystkich najniższych agregacji przyjętych w POPT II. Ponieważ udział wyodrębnionych sekcji PKD w poszczególnych województwach jest inny, operacja ta różnicuje województwa pod względem proporcji pracujących do zatrudnionych[22].

Kolejnym problemem związanym z czynnikiem pracy było oszacowanie wynagrodzenia pracy jako zatrudnionych do wynagrodzenia pracy jako pracujących. Tę dodatkową operację należało przeprowadzić we wszystkich rachunkach, w których za czynnik pracy przyjęto liczbę pracujących. Operacja ta była niezbędna do wyznaczenia udziału pracy w gospodarce, związanej z elastycznością α, która w zależności od tego, czy dotyczy zatrudnionych, czy pracujących, jest indeksowana jako α_z albo α_p. Posłużono się tutaj wzorem:

$$WP_{Psw} = \frac{H_{Ps}}{H_{Zs}}WP_{Zsw} \tag{23}$$

gdzie:
WP_{Psw} – obliczone wynagrodzenie pracy dla pracujących według sekcji PKD i województw,

[22] W poszczególnych sekcjach PKD ta proporcja jest względnie jednolita, także w ujęciu przestrzennym. Największe różnice pomiędzy liczbą pracujących a liczbą zatrudnionych obserwuje się dla: rolnictwa i usług związanych z turystyką, usług gastronomicznych, drobnych usług itp. Najmniejsze dla: administracji państwowej, przemysłu ciężkiego, sektorów opanowanych przez wielkie firmy, monopole, w bankowości itp.

H_{Ps} – liczba godzin przepracowanych przez pracujących dla sekcji (dane z TT),

H_{Zs} – liczba godzin przepracowanych przez zatrudnionych dla sekcji (dane z TT)[23],

WP_{Zsw} – wynagrodzenie pracy dla zatrudnionych, dla sekcji i województw (dane z BDL).

Ta operacja wymagała założenia, że samozatrudnieni wypłacają sobie w ramach nadwyżki operacyjnej netto, w tym dochodu mieszanego netto, takie same wynagrodzenie za swoją pracę, jakie zatrudnieni otrzymują za nią na podstawie umów o pracę za godzinę przepracowaną (wraz ze świadczeniami socjalnymi). Natomiast pozostała część dochodu mieszanego to zysk z kapitału, który należy uwzględnić w wynagrodzeniu kapitału[24]. Powyższe założenie nie sprawdza się jednak w krajach mniej rozwiniętych gospodarczo, o dużym znaczeniu rolnictwa w gospodarce, dla których konieczne jest zastosowanie innych metod szacowania wynagrodzenia czynnika pracy.

Wobec tego w przypadku Polski przyjęto rozwiązanie hybrydowe, czyli zastosowano wzór (23) dla wszystkich sekcji oprócz sekcji A (rolnictwo), dla której wykorzystano wzór:

$$WP_{Psw} = \left(WP_{Zs} + \frac{DM_s \times WP_{Zs}}{WDB_s - DM_s}\right)\frac{WDB_{sw}}{WDB_s} \tag{24}$$

Jest on oparty na koncepcji, że do wynagrodzenia pracy dla zatrudnionych WP_{Zs} na poziomie danej sekcji (w tym przypadku rolnictwa) należy dodać pewną część nadwyżki operacyjnej netto wraz z dochodem mieszanym netto DM_s na poziomie sekcji. Zakłada się, że ta nadwyżka jest wraz z dochodem mieszanym dzielona pomiędzy pracę a kapitał w takiej samej proporcji, jak reszta dochodu czynników (czyli $WDB_s - DM_s$) w pozostałej gospodarce na poziomie sekcji. Dane pochodzące z TT należało oszacować przy użyciu struktury pozyskanej z BDL, bazującej na proporcji pomiędzy wartością dodaną brutto dla sekcji i województw WDB_{sw}, a wartością dodaną brutto dla sekcji WDB_s[25].

[23] W teorii najodpowiedniejszą miarą czynnika pracy jest liczba godzin przepracowanych, dlatego zastosowano właśnie tę miarę, wykorzystując dostępność odpowiednich danych.

[24] Jest to wariant jednego z trzech sposobów doszacowania udziału pracy w wynagrodzeniu czynników zob. ILO (2014, s. 173). Sposób ten jest przyjęty także w innych rachunkach produktywności (OECD, 2001, s. 39 i 45), co ma swoje konsekwencje (OECD, 2001, s. 47).

[25] Metoda ta stanowi rozwinięcie jednej z koncepcji zaprezentowanych w wersji uproszczonej w ILO (2014, s. 173). W innych sekcjach niekiedy także nie wykorzystuje się pracowników najemnych, np. w sekcji G – Handel hurtowy i detaliczny, ale w zasadzie tylko w sekcji A powstają sprzeczności w postaci ujemnych wartości wynikowych rezydualnie obliczanych dla kapitału, czego skutkiem są problemy metodologiczne i rachunkowe.

KORZYŚCI ANALITYCZNE Z WYKONANIA DEKOMPOZYCJI W CZTERECH WARIANTACH

Dekompozycje czynnikowe wykonane w ramach POPT I i POPT II mają istotną przewagę nad dotychczas przeprowadzonymi w Polsce rachunkami produktywności gospodarki KLEMS, które dotyczą uwzględnienia wymiaru przestrzennego. O ile rachunek KLEMS jest realizowany obecnie tylko dla gospodarki całego kraju, o tyle rachunki w ramach POPT I i POPT II są wykonane także dla poszczególnych województw. Ponadto zostały one również wykonane w przeliczeniu na osobę biorącą udział w procesie produkcyjnym oraz dla odchyleń od średniej, co wskazano w części metodologicznej. Obecnie przeprowadzenie rachunku KLEMS na poziomie województw nie jest jeszcze możliwe z powodu trwania prac nad odpowiednim przygotowaniem danych. Rachunek KLEMS jest bardziej wymagający w zakresie danych wejściowych od dekompozycji zrealizowanych w ramach POPT I i POPT II. Dlatego wartość analityczna wykonanych rachunków dekompozycji polega przede wszystkim na możliwości rozszerzenia neoklasycznej analizy czynnikowej wzrostu gospodarczego o wymiar przestrzenny. W tym zakresie zostanie potwierdzona hipoteza badawcza, wskazana na początku artykułu, zgodnie z którą przeprowadzenie dekompozycji w czterech zaprezentowanych wariantach ma dużą wartość analityczną.

Na wykr. 1A uszeregowano województwa w kolejności od najmniejszego do największego tempa skumulowanego przyrostu względnego WDB w okresie 2004–2009. Jako punkt odniesienia dodano skumulowany przyrost względny WDB dla całego kraju[26]. Każdej agregacji na wykresie odpowiadają dwa słupki – lewy odnosi się do dekompozycji przyrostu WDB na kontrybucje wynagrodzeń czynników, prawy na kontrybucje zasobów czynników i TFP. Dla gospodarki polskiej jako całości słupki wyróżniono innymi odcieniami. Słupki umieszczone na wykresie na prawo od słupków dla Polski dotyczą województw rozwijających się w okresie 2004–2009 szybciej od średniej dla Polski. Na wykr. 1B, dotyczącym okresu 2010–2015, przyjęto identyczne założenia, zachowując kolejność uszeregowania województw z wykr. 1A. W badaniu zachowano pełną porównywalność dla ww. okresów sześcioletnich. Wykr. 1 dotyczy pracujących.

Jak widać na wykr. 1A (słupki prawe), kontrybucja TFP do wzrostu WDB była największa ze wszystkich wyróżnionych kontrybucji czynnikowych w okresie 2004–2009. W przypadku niektórych województw, np. dolnośląskiego, była na-

[26] Dane do obliczeń można znaleźć w: Kotlewski (2018b). Przyrosty skumulowane dla sześcioletnich okresów policzono według wzoru: $\Delta X_{skum} = (1 + \Delta X_{t=1})(1 + \Delta X_{t=2})(1 + \Delta X_{t=3})(1 + \Delta X_{t=4})(1 + \Delta X_{t=5})(1 + \Delta X_{t=6}) - 1$. Skumulowane TFP policzono rezydualnie jako różnicę pomiędzy skumulowanym przyrostem WDB a skumulowanymi kontrybucjami czynników produkcji.

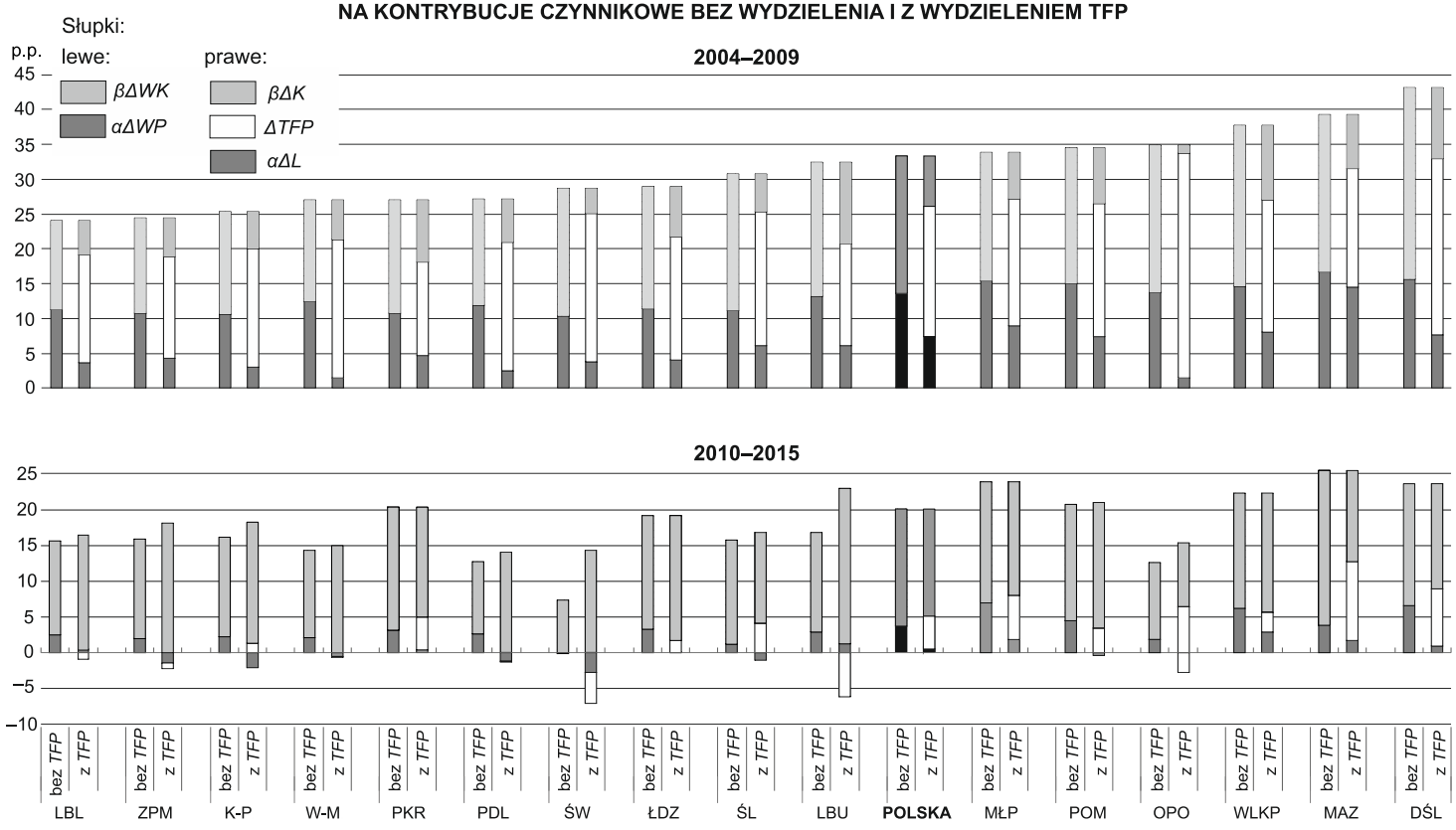
wet większa od wszystkich pozostałych wyróżnionych kontrybucji łącznie. Takie wyniki uzyskano także dla Polski. Wykonanie dekompozycji również na kontrybucje wynagrodzeń czynników (słupki lewe) pozwala stwierdzić, jaka część kontrybucji TFP przekształca się w wynagrodzenie pracy, a jaka – w wynagrodzenie kapitału. Przedstawione wykresy wskazują, że w okresie 2004–2009 przyrost TFP skutkował w większym stopniu przyrostem wynagrodzenia kapitału niż przyrostem wynagrodzenia pracy w większości województw (z wyjątkiem m.in. warmińsko-mazurskiego) oraz dla całego kraju. To oznacza, że w tym okresie produktywność kapitału rosła szybciej niż produktywność pracy (przy neoklasycznym założeniu o wynagrodzeniu czynników zgodnie z ich krańcowymi produktywnościami). Dekompozycja w układzie województw pozwala zaobserwować dystrybucję przestrzenną wpływu TFP na przyrosty wynagrodzeń czynników.

W okresie 2010–2015 gospodarka Polski znacznie się zmieniła, przede wszystkim zwolniła – skumulowany wzrost w okresie 2004–2009 wyniósł ok. 30%, podczas gdy w okresie 2010–2015 było to ok. 20%. Zmianom uległ ranking najszybciej rozwijających się województw – liderem nie jest już dolnośląskie, lecz mazowieckie, wystąpiły również inne zmiany w rankingu. Kontrybucja TFP w skali całej gospodarki kraju przestała być najważniejszą kontrybucją do wzrostu WDB[27] (spośród wyróżnionych kontrybucji w omawianym rachunku dekompozycji). Spadek kontrybucji TFP spowodował, że w warunkach jego zróżnicowania pomiędzy województwami wystąpiły przypadki ujemnej kontrybucji TFP dla niektórych województw[28]. Co więcej, o ile w okresie 2004–2009 większość przyrostu produktywności TFP przekształcała się w przyrost wynagrodzenia kapitału, o tyle w okresie 2010–2015 ten przyrost produktywności w zdecydowanej większości związany był z przyrostem produktywności pracy (na podstawie założeń neoklasycznych) z wyjątkiem tylko kilku województw, w tym woj. mazowieckiego, w których nadal przeważa wzrost produktywności kapitału nad wzrostem produktywności pracy. Spadek przyrostu produktywności kapitału jest zgodny z trendem światowym zaobserwowanym przez Acemoglu (2003), a także przez Klumpa, McAdama i Willmana (2004). Temu spadkowi przyrostu produktywności kapitału towarzyszył jednak wzrost jego kontrybucji bezpośredniej (tj. niezwiązanej z TFP) oraz kontrybucji jego wynagrodzenia na tle pozostałych wyróżnionych kontrybucji, nad którymi uzyskał pełną przewagę ilościową (słupki dla kapitału na wykr. 1B są największe).

[27] Spadek znaczenia TFP znajduje potwierdzenie w innych badaniach (zob. Gradzewicz, Growiec, Kolasa, Postek i Strzelecki, 2014, s. 23) oraz Kotlewski i Błażej (2016, 2018).

[28] Badania te wskazują również, że w niektórych latach wystąpiły przypadki ujemnej kontrybucji TFP dla całego kraju.

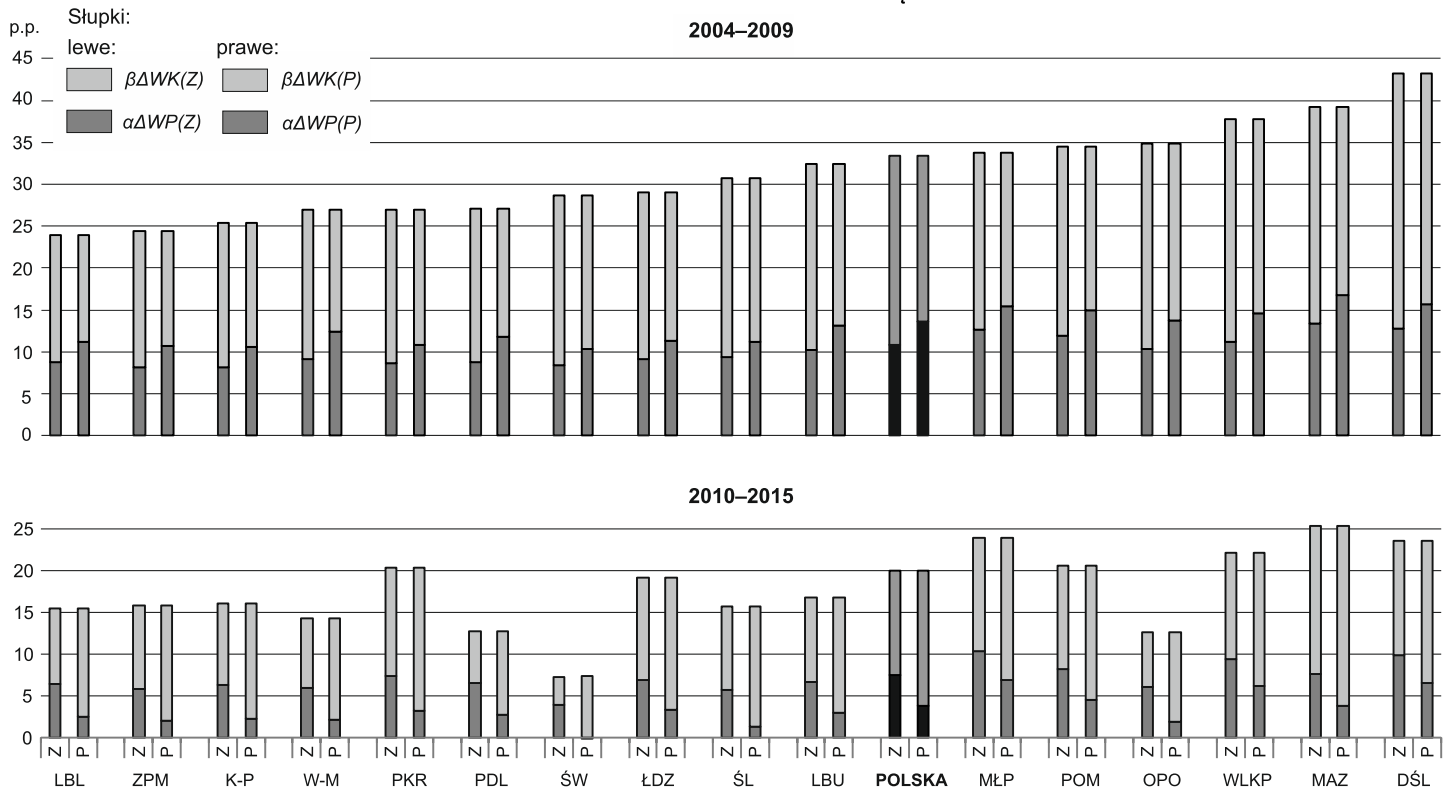
WYKR. 1. PORÓWNANIE DEKOMPOZYCJI SKUMULOWANYCH PRZYROSTÓW WDB NA KONTRYBUCJE CZYNNIKOWE BEZ WYDZIELENIA I Z WYDZIELENIEM TFP



U w a g a. Kontrybucja: *αΔWP* – wynagrodzenia pracy, *βΔWK* – wynagrodzenia kapitału, *αΔL* – zasobu pracy, *βΔK* – zasobu kapitału, *ΔTFP* – reszty Solowa. Województwa: LBL – lubelskie, ZPM – zachodniopomorskie, K-P – kujawsko-pomorskie, W-M – warmińsko-mazurskie, PKR – podkarpackie, PDL – podlaskie, ŚW – świętokrzyskie, ŁDZ – łódzkie, ŚL – śląskie, LBU – lubuskie, MŁP – małopolskie, POM – pomorskie, OPO – opolskie, WLKP – wielkopolskie, MAZ – mazowieckie, DŚL – dolnośląskie.

Ź r ó d ł o: opracowanie własne na podstawie danych GUS.

WYKR. 2. PORÓWNANIE DEKOMPOZYCJI SKUMULOWANYCH PRZYROSTÓW WDB NA KONTRYBUCJE WYNAGRODZEŃ CZYNNIKÓW DLA ZATRUDNIONYCH I DLA PRACUJĄCYCH



U w a g a. Kontrybucja wynagrodzenia: *αΔWP(Z)* – pracy dla zatrudnionych, *βΔWK(Z)* – kapitału dla zatrudnionych, *αΔWP(P)* – pracy dla pracujących, *βΔWK(P)* – kapitału dla pracujących. Oznaczenie województw jak przy wykr. 1.

Ź r ó d ł o: jak przy wykr. 1.

Ujemna kontrybucja TFP wystąpiła szczególnie wyraźnie w przypadku województw świętokrzyskiego i lubuskiego. Przyrost zasobu kapitału był w nich dużo większy niż przyrost wynagrodzenia kapitału, czyli w okresie 2010–2015 produktywność kapitału (na podstawie założeń neoklasycznych) spadała. Obniżenie produktywności kapitału można wiązać z ujemną kontrybucją TFP, na co wskazują białe słupki na wykr. 1B dla tych województw. W latach 2004–2009, w przeciwieństwie do lat 2010–2015, tempo przyrostu wynagrodzenia kapitału było wyższe od tempa przyrostu samego kapitału dla wszystkich województw. Wszystkie te analizy są możliwe do przeprowadzenia w ujęciu województw tylko wtedy, gdy jednocześnie wykonywana jest dekompozycja przyrostu WDB obejmująca kontrybucje wynagrodzeń czynników oraz odrębnie zasobów czynników i TFP, jak w omawianym rachunku. Wykres 1 można wykonać dla każdej sekcji lub grupy sekcji uwzględnionej w badaniu, pogłębiając zatem te analizy[29]. W tym układzie niemal identyczne wyniki jak dla pracujących uzyskano dla zatrudnionych, stąd nie przedstawiono ich na odrębnym wykresie.

Wartość analityczną wykonanych dekompozycji, zarówno dla zatrudnionych, jak i dla pracujących, przedstawiono na wykr. 2. Przyjęto takie same założenia jak dla wykr. 1, przy czym wszystkie słupki, zarówno lewe, jak i prawe, dotyczą dekompozycji przyrostu WDB na kontrybucje wynagrodzeń czynników. Słupek lewy odpowiada czynnikowi pracy rozumianemu jako zatrudnieni, a słupek prawy – czynnikowi pracy rozumianemu jako pracujący. W okresie 2004–2009 dla wszystkich województw, a zatem i dla kraju łącznie, kontrybucja wynagrodzenia pracujących była większa od kontrybucji wynagrodzenia zatrudnionych do przyrostu WDB. Sytuacja ta jest częściowo wynikiem przyjętego parametru wagi, czyli udziału pracy oznaczanego jako α (z odpowiednimi indeksami dla każdego poziomu agregacji i każdego okresu), który większy był dla pracujących[30]. Przedmiotem ewentualnej szczegółowej analizy może być rozkład tej różnicy pomiędzy województwami.

Nie mniej interesujące wyniki uzyskuje się dla okresu 2010–2015, szczególnie w porównaniu z latami 2004–2009. Kontrybucja wynagrodzenia pracujących do przyrostu WDB w tym okresie była mniejsza od kontrybucji wynagrodzenia zatrudnionych, co wskazuje na ujemną kontrybucję wynagrodzenia samozatrudnionych do przyrostu WDB w latach 2010–2015. Nastąpiło zatem odwrócenie sytuacji z okresu 2004–2009, kiedy samozatrudnieni istotnie przyczyniali się do przyrostu WDB. Podobne wyniki można zaobserwować także na wykresach dla zatrudnionych i pracujących dla dekompozycji z uwzględnieniem TFP, ale są one mniej wyraziste, dlatego ich nie przedstawiono.

[29] Oznacza to powielenie wykr. 1 12 razy, a także możliwość opracowania wszystkich tych wykresów dla zatrudnionych. Podobnie rzecz się ma z wykr. 2, który też można powielić 12 razy, jak również wykonać alternatywne wersje wykresów (z uwzględnieniem w słupkach lewych i prawych TFP).

[30] Udział zatrudnionych wraz z samozatrudnieniem, czyli pracujących, jest w WDB większy od udziału samych zatrudnionych w WDB.

W okresie 2004–2009 samozatrudnieni pełnili ważną funkcję we wzroście gospodarczym, ponieważ rozwijały się wtedy stosunkowo dobrze małe, często jednoosobowe firmy, przede wszystkim w sektorze rolnym oraz usługach, w szczególności turystycznych, gastronomicznych, prawnych itp. Różnica pomiędzy kontrybucją wynagrodzenia pracujących a kontrybucją wynagrodzenia zatrudnionych jest szczególnie duża w przypadku województw wielkopolskiego (w którym rolnictwo ma duże znaczenie) oraz warmińsko-mazurskiego (o istotnej roli turystyki i małej gastronomii). Różnica ta jest większa również dla woj. mazowieckiego (z dużym udziałem usług prawnych, małego pośrednictwa finansowego itp.) w stosunku do dolnośląskiego. W latach 2010–2015 nastąpiła konsolidacja tych usług lub przejęcie ich przez większe i bardziej wyspecjalizowane firmy, czemu towarzyszyło zmniejszenie „kominów" wynagrodzeń dla samozatrudnionych z powodu wzrostu konkurencji na rynku upodabniającym się pod względem płynności do rynków starych zachodnich gospodarek.

Wykresy 1 i 2 potwierdzają wzrost znaczenia kontrybucji kapitału i jego wynagrodzenia do przyrostu WDB w okresie 2010–2015. Ta obserwacja jest zgodna z generalnymi spostrzeżeniami dotyczącymi rynków wschodzących (van Ark, 2016). W przypadku Polski można to częściowo wiązać z dostępem do funduszy Unii Europejskiej, które stymulują wzrost inwestycji, np. infrastrukturalnych, przekształcających się efektywnie we wzrost gospodarczy dopiero w długim okresie. Kontrybucje zasobu kapitału oraz wynagrodzenia kapitału (w szczególności te pierwsze) powinny być zatem powiększone w okresie 2010–2015 w stosunku do stanu z okresu 2004–2009. Wszystkie te obserwacje stają się pełniejsze[31] dzięki wykonaniu dekompozycji według dwóch dychotomii, tj. zarówno na kontrybucje wynagrodzeń czynników oraz kontrybucje ich zasobów i TFP, jak i przy uwzględnieniu zatrudnionych oraz pracujących.

PODSUMOWANIE

W artykule przedstawiono rachunek dekompozycji czynnikowej zrealizowany w czterech wariantach. Uwzględniono w nim kontrybucje wynagrodzeń czynników produkcji lub kontrybucje ich zasobów i TFP oraz zrealizowano go zarówno dla zatrudnionych, jak i dla pracujących. Wykonano go w podziale na sekcje lub grupy sekcji dostępnych w zasobach GUS oraz na województwa. Rachunek ten wymagał odpowiedniego przygotowania danych wejściowych, czego dokonano za pomocą autorskiej metody na podstawie dotychczasowych badań statystycznych.

Istotna jest przy tym idea dekompozycji wielowariantowej, która według wiedzy autora stanowi zupełnie nowe podejście do zagadnienia dekompozycji wzrostu gospodarczego. Przeprowadzona w artykule analiza potwierdziła, że

[31] Inne obserwacje tego rodzaju mają bardziej cząstkowy charakter.

dość pracochłonne operacje związane z wykonaniem odpowiednich obliczeń są pomimo ich złożoności użyteczne, ponieważ umożliwiają uzyskanie dodatkowych informacji o stanie gospodarki, jednocześnie w aspekcie regionalnym i sektorowym.

Wartość analityczna przeprowadzonych rachunków przesądza o ich znaczeniu dla badaczy wzrostu gospodarczego. Dzięki tego rodzaju rachunkom analiza wzrostu gospodarczego, szczególnie w aspekcie regionalnym, może zostać znacznie pogłębiona, szczególnie gdy oprócz aspektu regionalnego uwzględni się jeszcze aspekt sektorowy oraz oba te aspekty jednocześnie[32]. Dlatego pożądane byłoby zapoznanie się z wynikami omawianego badania w ramach edycji POPT II przez odpowiednie organy publiczne i firmy analityczne.

BIBLIOGRAFIA

Acemoglu, D. (2003). Labor- and Capital-Augmenting Technical Change. *Journal of the European Economic Association*, *1*(1), 1–37. DOI: 10.1162/154247603322256756.

van Ark, B. (2016). *Are Emerging Markets Still Emerging?* Pobrane z: http://www.worldklems.net//conferences/worldklems2016/worldklems2016_van-Ark_slides_2.pdf.

Gradzewicz, M., Growiec, J., Kolasa, M., Postek, Ł., Strzelecki, P. (2014). *Poland's exceptional performance during the world economic crisis: New growth accounting evidence* (NBP Working Paper No. 186). Pobrane z: https://pdfs.semanticscholar.org/672f/3d78a635eee6ea1b3967577a72a01b49d473.pdf.

Havik, K., Mc Morrow, K., Orlandi, F., Planas, C., Raciborski, R., Röger, W., Rossi, A., Thum-Thysen, A., Vandermeulen, V. (2014). The Production Function Methodology for Calculating Potential Growth Rates & Output Gaps. *European Commission, Economic Papers*, *535*, 1–120. DOI: 10.2765/71437.

Hulten, C. R. (2009). *Growth Accounting* (NBER Working Paper No. 15341). Pobrane z: https://www.nber.org/papers/w15341.pdf.

ILO (2014). *World of Work Report 2014.* Genève: International Labour Organization.

Jorgenson, D. W. (1963). Capital Theory and Investment Behavior. *The American Economic Review*, *53*(2), 247–259.

Jorgenson, D. W. (1989). Productivity and Economic Growth. W: R. E. Berndt, E. J. Triplett (red.), *Fifty Years of Economic Measurement: The Jubilee of the Conference on Research in Income and Wealth* (s. 19–118). Chicago: The University of Chicago Press.

Jorgenson, D. W., Gollop, F. M., Fraumeni, B. M. (1987). *Productivity and US Economic Growth*. Cambridge, Massachusetts: Harvard University Press.

Jorgenson, D. W., Griliches, Z. (1967). The explanation of Productivity Change. *The Review of Economic Studies*, *34*(3), 249–283. DOI: 10.2307/2296675.

Jorgenson, D. W., Ho, M. S., Stiroh, K. J. (2005). *Information Technology and the American Growth Resurgence*. Cambridge, Massachusetts: The MIT Press.

Klump, R., McAdam, P., Willman, A. (2004). *Factor Substitution and Factor Augmenting Technical Progress in the US: A Normalized Supply-Side System Approach* (ECB Working Paper Series

[32] I czego efektem mogłoby być odpowiednie opracowanie książkowe, o ile pojawi się popyt na tego rodzaju wiedzę.

No. 367). Pobrane z: https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp367.pdf?1736264772425dc185f0c26ad9f93ebd.

Kotlewski, D. (2015). Część B: Dekompozycje czynnikowe WDB na zatrudnionego. W: M. Lewandowski, *Metoda dekompozycji Produktu Krajowego Brutto (PKB) oraz Wartości Dodanej Brutto (WDB) w zastosowaniu do analizy struktury różnic regionalnych*. Pobrane z: http://stat.gov.pl/statystyka-regionalna/statystyka-dla-polityki-spojnosci/statystyka-dla-polityki-spojnosci-2013-2015/badania/dezagregacja-wskaznikow-z-obszaru-rachunkow-narodowych-i-regionalnych/.

Kotlewski, D. (2017a). Dekompozycje wartości dodanej brutto na wkłady wynagrodzeń czynników praca i kapitał. *Wiadomości Statystyczne*, *2*(669), 31–51.

Kotlewski, D. (2017b). Problem cen w regionalnym rachunku produktywności. *Wiadomości Statystyczne*, *12*(679), 50–63.

Kotlewski, D. (2018a). *Rozdział I – Raport metodologiczny, Część B; Rozdział IV – Raport analityczny, Część B*. W: M. Lewandowski, *Identyfikacja źródeł zróżnicowania regionalnego Polski przy wykorzystaniu metod dekompozycji wzrostu i różnic Produktu Krajowego Brutto (PKB) oraz Wartości Dodanej Brutto (WDB) per capita*. Pobrane z: http://stat.gov.pl/statystyka-regionalna/statystyka-dla-polityki-spojnosci/statystyka-dla-polityki-spojnosci-2016-2018/badania/ekonomia/.

Kotlewski, D. (2018b). *Załączniki*. Pobrane z: https://stat.gov.pl/statystyka-regionalna/statystyka-dla-polityki-spojnosci/statystyka-dla-polityki-spojnosci-2016-2018/badania/ekonomia/.

Kotlewski, D., Błażej, M. (2016). Metodologia rachunku produktywności KLEMS i jego implementacja w warunkach polskich. *Wiadomości Statystyczne*, *9*(664), 86–108.

Kotlewski, D., Błażej, M. (2018). Implementation of KLEMS Economic Productivity Accounts in Poland. *Acta Universitatis Lodziensis. Folia Oeconomica*, *2*(334), 7–18. DOI: 10.18778/0208-6018.334.01.

Lewandowski, M. (2015). *Metoda dekompozycji Produktu Krajowego Brutto (PKB) oraz Wartości Dodanej Brutto (WDB) w zastosowaniu do analizy struktury różnic regionalnych.* Pobrane z: http://stat.gov.pl/statystyka-regionalna/statystyka-dla-polityki-spojnosci/statystyka-dla-polityki-spojnosci-2013-2015/badania/dezagregacja-wskaznikow-z-obszaru-rachunkow-narodowych-i-regionalnych/.

Milana, C. (2009). *Solving the Index-Number Problem in a Historical Perspective* (EU KLEMS Working Paper No. 43). Pobrane z: http://www.euklems.net/pub/no43(online).pdf.

O'Mahony, M., Timmer, M. P. (2009). Output, Input and Productivity Measures at the Industry Level: The EU KLEMS Database. *The Economic Journal, 119*(538) F374–F403. DOI: 10.1111/j.1468-0297.2009.02280.x.

OECD. (2001). *Measuring Productivity.* Paris: OECD.

Solow, R. M. (1956). A Contribution to the Theory of Economic Growth. *The Quarterly Journal of Economics, 70*(1), 65–94. DOI: 10.2307/1884513.

Solow, R. M. (1957). Technical Change and the Aggregate Production Function. *Review of Economics and Statistics*, *39*(3), 312–320.

Timmer, M., van Moergastel, T., Stuivenwold, E., Ypma, G., O'Mahony, M., Kangasniemi, M. (2007). *EU KLEMS Growth and Productivity Accounts. Version 1.0. PART I Metodology*. Birmingham: EU KLEMS Consortium. Pobrane z: http://www.euklems.net/data/EUKLEMS_Growth_and_Productivity_Accounts_Part_I_Methodology.pdf.

Wiadomości Statystyczne. The Polish Statistician, 2019, vol. 64, 9, 60–74 DOI: 10.5604/01.3001.0013.7590
Studia interdyscyplinarne. Wyzwania badawcze / Interdisciplinary studies. Research challenges

Pozyskiwanie i analiza danych na temat ofert pracy z wykorzystaniem big data

Jacek Maślankowski[a]

Streszczenie. Celem artykułu jest zaprezentowanie korzyści wynikających z wykorzystania na potrzeby statystyki publicznej (rynku pracy) narzędzi do automatycznego pobierania danych na temat ofert pracy zamieszczanych na stronach internetowych zaliczanych do zbiorów big data, a także związanych z tym wyzwań. Przedstawiono wyniki eksperymentalnych badań z wykorzystaniem metod web scrapingu oraz text miningu. Analizie poddano dane z lat 2017 i 2018 pochodzące z najpopularniejszych portali z ofertami pracy. Odwołano się do danych Głównego Urzędu Statystycznego (GUS) zbieranych na podstawie sprawozdania Z-05. Przeprowadzona analiza prowadzi do wniosku, że web scraping może być stosowany w statystyce publicznej do pozyskiwania danych statystycznych z alternatywnych źródeł, uzupełniających istniejące bazy danych statystycznych, pod warunkiem zachowania spójności z istniejącymi badaniami.
Słowa kluczowe: big data, text mining, web scraping, rynek pracy

The collection and analysis of the data on job advertisements with the use of big data

Summary. The goal of this paper is to present, on the one hand, the benefits for official statistics (labour market) resulting from the use of web scraping methods to gather data on job advertisements from websites belonging to big data compilations, and on the other, the challenges connected to this process. The paper introduces the results of experimental research where web-scraping and text-mining methods were adopted. The analysis was based on the data from 2017–2018 obtained from the most popular job-searching websites, which was then collated with Statistics Poland's data obtained from Z-05 forms. The above-mentioned analysis demonstrated that web-scraping methods can be adopted by public statistics services to obtain statistical data from alternative sources complementing the already-existing databases, providing the findings of such research remain coherent with the results of the already-existing studies.
Keywords: big data, text mining, web scraping, labour market

JEL: C18, M15

[a] Uniwersytet Gdański, Wydział Zarządzania.

Wraz z rozwojem pakietów komputerowych i narzędzi statystycznych pojawiły się nowe możliwości w zakresie pozyskiwania danych dla statystyki publicznej. Zalicza się do nich narzędzia big data. Należy zaznaczyć, że termin *big data* nie oznacza technologii. Niekiedy odnosi się do dużego zbioru danych, jednak w tym artykule występuje jako narzędzia pozwalające na przechowywanie i przetwarzanie danych, co nie byłoby możliwe przy wykorzystaniu tradycyjnych metod, np. relacyjnych baz danych (Shahin, 2016). Najbardziej znana definicja obejmuje takie cechy zbioru danych typu big data, jak duża ilość (ang. *volum*e), duża zmienność (ang. *velocity*) oraz duża różnorodność (ang. *variety*). Została ona zaproponowana przez pracowników firmy Gartner w raporcie opublikowanym na blogu firmy META Group (Douglas, 2001). Często podawane są też atrybuty dotyczące wiarygodności (ang. *veracity*) oraz wartości danych (ang. *value*). Oznacza to, że dane ze zbiorów big data powinny charakteryzować się wysoką wiarygodnością, rozumianą również jako wysoka jakość, a także stanowić wartość dla użytkowników.

Jedną z metod wykorzystywanych podczas gromadzenia danych typu big data jest automatyczne pobieranie danych z internetu, znane powszechnie jako web scraping. W tym celu wykorzystuje się programy zwane robotami internetowymi, których zadaniem jest pozyskiwanie odpowiednich informacji, w niniejszym artykule – dotyczących ofert pracy. Dodatkowo, aby uzyskać informacje na temat zawodu, wykształcenia czy kwalifikacji, zastosowano metodę text miningu, która służy do eksploracji i obróbki tekstu[1].

Celem artykułu jest zaprezentowanie korzyści wynikających z wykorzystania na potrzeby statystyki publicznej (rynku pracy) narzędzi do automatycznego pobierania danych na temat ofert pracy zamieszczanych na stronach internetowych, zaliczanych do zbiorów typu big data, a także związanych z tym wyzwań. Przedstawiono wyniki eksperymentalnych badań z wykorzystaniem metod web scrapingu oraz text miningu.

BIG DATA JAKO ALTERNATYWNE ŹRÓDŁO DANYCH DLA STATYSTYKI PUBLICZNEJ

Zastosowaniu big data w statystyce publicznej poświęcono wiele artykułów naukowych. W literaturze podkreśla się przede wszystkim wykorzystanie tego typu zbiorów danych do obliczania wskaźnika cen towarów i usług konsumpcyjnych (ang. CPI – Consumer Price Index). Według Hackla (2016) w 2016 r. 17 krajów europejskich pracowało nad rozwiązaniami dotyczącymi big data w celu dostarczania danych na potrzeby tego wskaźnika. Listę alternatywnych źródeł danych i możliwości ich zastosowania w statystyce publicznej prezentuje zestawienie 1.

[1] https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/index.php/Main_Page (dostęp: 30.06.2018).

ZESTAWIENIE 1. ALTERNATYWNE ŹRÓDŁA DANYCH I ICH ROLA W STATYSTYCE PUBLICZNEJ

Źródła danych	Obszar potencjalnego wykorzystania
Skanery kodów kreskowych	statystyka cen, statystyki ekonomiczne
Lokalizacja telefonów komórkowych	statystyka turystyki, statystyka ludności i migracji
Sensory drogowe	statystyka transportu
Mierniki zużycia energii	statystyka ludności, statystyka gospodarstw domowych
Zdjęcia satelitarne, zdalne sensory	statystyka rolnictwa, leśnictwa, rybołówstwa, statystyka środowiska naturalnego
Serwisy społecznościowe, internet	statystyka rynku pracy, statystyka ludności i migracji, statystyka dochodów i konsumpcji gospodarstw domowych, statystyka cen, statystyka zdrowia, statystyka społeczna
Strony WWW z ofertami pracy	statystyka rynku pracy
Ruch samolotów	statystyka transportu, statystyka ochrony środowiska
Strony WWW: nieruchomości, działalność e-commerce	statystyka cen

Ź r ó d ł o: opracowanie własne na podstawie: Hackl, 2016; Kitchin, 2015.

Inne źródła danych dostarczają informacji o ruchu pojazdów, jak również dotyczą przetwarzania treści zamieszczanych w serwisach społecznościowych (Daas, Puts, Buelens i Hurk, 2015). Nazywa się je często niestatystycznymi, czyli nieutworzonymi na potrzeby statystyki publicznej (Beręsewicz i Szymkowiak, 2015), lub pozastatystycznymi. Dane z takich źródeł mają szerokie zastosowanie w statystyce publicznej. Przykładowo w ramach grantu Komisji Europejskiej ESSNet Big Data prowadzone są prace eksperymentalne, które bazują na danych pobieranych ze stron internetowych przedsiębiorstw i mają na celu identyfikację obecności przedsiębiorstwa w mediach społecznościowych czy też działalności e-commerce (wsparcie dla europejskiego badania ICT in Enterprises – ICT w przedsiębiorstwach) oraz weryfikację rodzaju działalności pod kątem zgodności z kodem w Polskiej Klasyfikacji Działalności (PKD) lub, w szerszym ujęciu, w europejskiej klasyfikacji NACE[2].

Wykorzystanie alternatywnych źródeł danych typu big data w statystyce publicznej może skutkować zmianą modelu funkcjonowania niektórych obszarów statystyki. Zgodnie z wieloletnią tradycją statystyka publiczna bazuje na danych, które pozyskuje za pomocą kwestionariuszy i sprawozdań (Braaksma i Zeelenberg, 2015). W ciągu ostatnich lat obserwuje się stopniowe odchodzenie od tradycyjnych sprawozdań na rzecz administracyjnych źródeł danych. Ma to miejsce w takich dziedzinach, jak rynek pracy czy edukacja, w której nastąpiło przejście ze sprawozdań dla szkół na System Informacji Oświatowej czy Zintegrowany System Informacji o Nauce i Szkolnictwie Wyższym POL-on[3]. Wykorzystanie

[2] https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/index.php/Main_Page (dostęp: 30.06.2018).

[3] Zob. program badań statystycznych statystyki publicznej na rok 2018, http://bip.stat.gov.pl/dzialalnosc-statystyki-publicznej/program-badan-statystycznych/pbssp-2018 (dostęp: 25.02.2019).

big data spowoduje większy nacisk na metodologię oraz sposoby analizy danych zamiast skupiania się na przygotowywaniu kwestionariuszy oraz kontroli procesu sprawozdawczości. Wzrośnie tym samym zapotrzebowanie na inżynierów danych (ang. *data scientists*) mających wiedzę z pogranicza informatyki i statystyki, niezbędną do budowy repozytoriów i przetwarzania dużych zbiorów, a w szczególności integracji różnych zbiorów danych, takich jak bazy relacyjne i typowe dla big data bazy NoSQL (Miller, 2014). Ta zmiana to jeden z elementów modernizacji statystyki publicznej, w ramach której powinny zostać wypracowane wytyczne dotyczące jakości danych, partnerstwa z firmami prywatnymi je dostarczającymi, aspektów prywatności danych, stosowanej metodologii oraz wykorzystywanych technologii (Vale, 2015). Wówczas statystyka publiczna mogłaby przyjąć dla wybranych dziedzin strategię rozwoju sterowaną danymi (ang. *data driven approach*), która może wymagać sformułowania metod gromadzenia, współdzielenia oraz ponownego wykorzystywania dużych zbiorów danych (Rousidis, Garoufallou, Balatsoukas i Sicilia, 2014). Należy jednak mieć na uwadze, że administracyjne źródła danych oraz zbiory typu big data nie zawierają wszystkich niezbędnych informacji, na które jest zapotrzebowanie użytkowników danych statystycznych, w tym uwzględnionych w aktach prawnych obligujących statystykę publiczną do ich prezentowania w ściśle określonych przekrojach i terminach.

Przyjmuje się, że w statystyce publicznej można wykorzystać źródła danych big data na pięć sposobów (Kitchin, 2015):

1. całościowe zastąpienie istniejących źródeł statystycznych (istniejące dane);
2. częściowe zastąpienie istniejących źródeł statystycznych (istniejące dane);
3. dostarczenie komplementarnej informacji statystycznej w tej samej dziedzinie statystyki, jednak z innej perspektywy (dodatkowe dane);
4. skorygowanie szacunków pochodzących z istniejących źródeł statystycznych (poprawione dane);
5. dostarczenie zupełnie nowej informacji statystycznej w danej dziedzinie (nowe alternatywne źródło danych).

Należy zwrócić uwagę, że badanie dotyczące ofert pracy jest tematem opracowań naukowych powstających od wielu lat (Gałecka-Burdziak i Pater, 2015; Kureková, Beblavý i Thum-Thysen, 2015), co pozwala na analizę porównawczą stosowanych podejść i metodologii. Badane są przede wszystkim informacje o wolnych miejscach pracy publikowane w internecie.

ZASADY POBIERANIA DANYCH ZE STRON INTERNETOWYCH

Dane zamieszczane na stronach internetowych podlegają ochronie, nie można ich kopiować m.in. w celu udostępniania osobom trzecim. Ponadto w przypadku publikowania baz danych na stronie internetowej obowiązuje Ustawa

z dnia 27 lipca 2001 r. o ochronie baz danych, zgodnie z którą „pobieranie danych oznacza stałe lub czasowe przejęcie lub przeniesienie całości lub istotnej, co do jakości lub ilości, części zawartości bazy danych na inny nośnik, bez względu na sposób lub formę tego przejęcia lub przeniesienia". Regulacje te determinują dobór narzędzi oraz możliwości analizy dostępnych baz danych. Jednocześnie „producent bazy danych udostępnionej publicznie w jakikolwiek sposób nie może zabronić użytkownikowi korzystającemu zgodnie z prawem z takiej bazy danych, pobierania lub wtórnego wykorzystania w jakimkolwiek celu nieistotnej, co do jakości lub ilości, części jej zawartości". Należy zauważyć, że bazy danych udostępniane na stronach internetowych mają w większości charakter publiczny.

Poza kwestiami prawnymi dotyczącymi możliwości przetwarzania danych należy również wziąć pod uwagę wykorzystanie pliku *robots.txt*, który powinien znajdować się na każdej stronie internetowej. Określa on, jakiego rodzaju działania robotów są dozwolone w danym serwisie. Przykładowo wpis „User-agent: * Disallow: /" oznacza, że nie jest dozwolone pobieranie danych przez inne roboty niż wskazane w tym pliku[4]. Zasady poruszania się po stronie internetowej oraz dozwolone foldery są określane przez właścicieli serwisu.

Jeżeli działania robotów na stronie internetowej są dozwolone, wówczas możliwa staje się analiza zawartości witryn w czasie rzeczywistym. Dzięki temu nie trzeba pobierać ani przechowywać potrzebnych stron internetowych. Dobrą praktyką jest też pobieranie danych w interwałach czasowych, np. kolejnych części strony co 10 sekund, aby nadmiernie nie przeciążać serwerów właścicieli serwisu. Skutkiem nadmiernego obciążenia serwerów może być zablokowanie robota przez właścicieli serwisu. Dane na stronach internetowych występują w postaci częściowo ustrukturyzowanej, tj. zapisane są za pomocą znaczników HTML, które umożliwiają nawigację na stronie. Przykładowo znacznik określa sekcję, a – pojedynczą linię strony. Skojarzenie znaczników z klasami stylów pozwala nawigować na stronie narzędziom do web scrapingu, czyli zautomatyzowanego pobierania wybranych treści ze stron internetowych.

OFERTY PRACY ONLINE – ANALIZA

Pierwsze eksperymentalne pobieranie danych ze stron internetowych zawierających oferty pracy, z myślą o wykorzystaniu utworzonego zbioru danych w polskiej statystyce publicznej, zostało przeprowadzone przez autora niniejszego artykułu w 2013 r., co pozwoliło na wstępne wskazanie problemów z jakością danych (Maślankowski, 2014). Powtarzanie tego procesu przez kolejne lata umożliwiło sformułowanie wielu wniosków dotyczących zastosowania tego

[4] http://www.robotstxt.org (dostęp: 29.01.2019).

rodzaju źródeł danych w statystyce publicznej. Odnoszą się one przede wszystkim do jakości danych rozumianej jako reprezentatywność oraz stabilność źródeł danych. Obecnie autor pobiera dane codziennie, a wynikowy zbiór danych z ofertami pracy potwierdza wnioski formułowane w latach ubiegłych. W ramach prowadzonych prac jednostką badania jest oferta pracy na terytorium Polski zamieszczona online.

W badaniu jako źródła internetowe zawierające oferty pracy wykorzystano portale: Praca.money.pl, GoWork.pl, Jobs.pl, Pracuj.pl, jak również portale branżowe BazaOgloszen.nauka.gov.pl i Nabory.kprm.gov.pl oraz portale Biuletynu Informacji Publicznej (BIP). Dane pobierano codziennie za pomocą robotów internetowych, tj. przygotowanego przez autora narzędzia do web scrapingu. W tej części artykułu posłużono się danymi z lat 2017 i 2018 według stanu na II kwartał. Odwołano się także do oficjalnych danych statystycznych zbieranych przez Główny Urząd Statystyczny (GUS) na podstawie sprawozdania Z-05 Badanie popytu na pracę (2019).

Istotne jest zdefiniowanie, jakiego rodzaju informacje można pozyskiwać ze źródeł danych big data, tj. stron internetowych zawierających oferty pracy. W tym celu odniesiono się do wybranych pozycji sprawozdania Z-05 (zestawienie 2), należy jednak podkreślić, że badanie to oferuje bardzo szeroki zakres danych, dopasowywanych do zmieniających się potrzeb odbiorców, uwarunkowanych również wytycznymi unijnymi (GUS, 2019).

ZESTAWIENIE 2. WYBRANE INFORMACJE DOTYCZĄCE POPYTU NA PRACĘ PUBLIKOWANE PRZEZ GUS ORAZ MOŻLIWOŚCI WYKORZYSTANIA ŹRÓDEŁ DANYCH BIG DATA

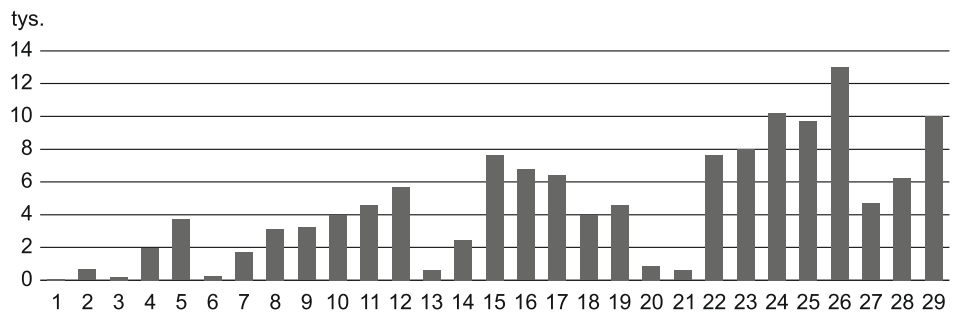
Rodzaje informacji objętej badaniem GUS	Uwagi dotyczące wykorzystania źródeł danych big data
Wolne miejsca pracy – w końcu kwartału	możliwe do pozyskania, brak reprezentacyjności i nieznana populacja
Wolne nowo utworzone miejsca pracy – w końcu kwartału	możliwe do obliczenia w długim szeregu czasowym, brak reprezentacyjności i nieznana populacja
Nowo utworzone miejsca pracy – w kwartale	trudne do uzyskania, brak reprezentacyjności i nieznana populacja
Zlikwidowane miejsca pracy – w kwartale	brak możliwości pozyskania danych

Ź r ó d ł o: opracowanie własne oraz GUS (2018).

Jak można wywnioskować z zestawienia 2, obecnie nie jest możliwe pozyskiwanie pełnej informacji na temat popytu na pracę, jak ma to miejsce w przypadku sprawozdania Z-05. Internetowe zbiory danych charakteryzują się nieznaną populacją, więc wyniki mogą znacznie odbiegać od oficjalnych danych publikowanych przez GUS.

Dalsza analiza danych dotyczy liczby ofert pracy pozyskiwanych z portali internetowych. Należy wziąć pod uwagę, że portale internetowe, na których publikowane są oferty pracy, mają zwykle wysokie standardy dotyczące aktualności zamieszczanych ofert – dane są usuwane najczęściej po upływie ok. miesiąca. Na wykr. 1 przedstawiono aktualne oferty pracy zamieszczane na wybranym portalu w kolejnych dniach maja 2017 r.

WYKR. 1. NOWE OFERTY PRACY OPUBLIKOWANE OD 1 DO 29 MAJA 2017 R.

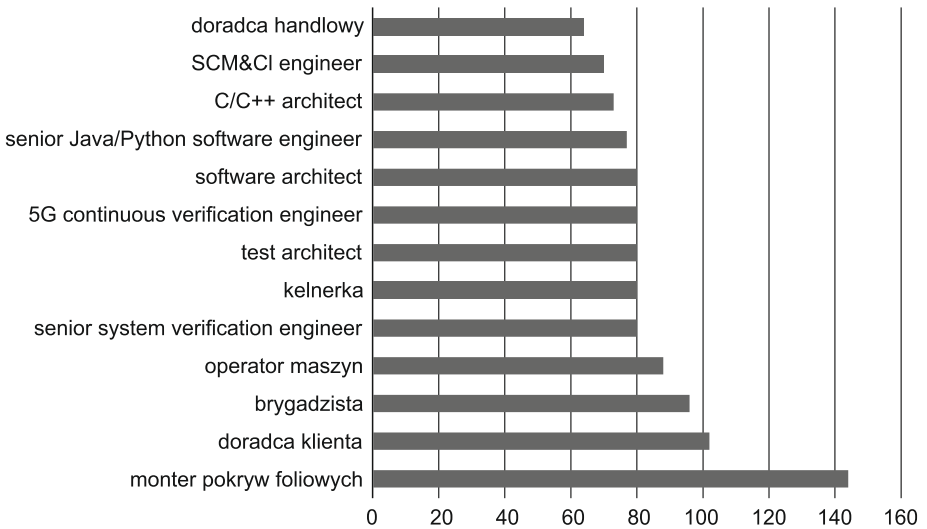


Ź r ó d ł o: opracowanie własne na podstawie wybranego portalu z ofertami pracy.

Wykres 1 przedstawia zasadę działania portali z ogłoszeniami o pracy, która polega na usuwaniu ofert pracy uznanych za nieaktualne. Uśredniając uzyskane wyniki, można zauważyć, że starszych ofert pracy jest znacząco mniej niż aktualnych. Na wykresie pominięto oferty sprzed maja 2017 r. Ich liczba ogółem wyniosła 325 (według daty publikacji oferty) w stosunku do 132575 ofert w maju 2017 r. Można również zaobserwować zjawisko polegające na codziennym umieszczaniu przez ogłoszeniodawców tych samych ofert pracy, aby były wyświetlane jako pierwsze przy domyślnym sortowaniu przeglądania ofert według daty dodania. Ten rodzaj aktywności firm na portalach z ogłoszeniami o pracy może występować kilka razy dziennie i skutkować błędami w obliczeniach. Skalę zjawiska prezentuje wykr. 2.

Wykres 2 dotyczy identycznych ofert, zawierających taki sam opis. Błędne wydaje się założenie, że firma poszukuje 80 czy 145 przedstawicieli danej profesji. Tym samym takie przykłady powinny być traktowane jako duplikaty. Aby wyeliminować tego typu problemy, opracowano metodę deduplikacji danych, która polega na wykrywaniu podobieństw w ofertach i usuwaniu duplikatów. Na podstawie analizy nazwy oferty, firmy, miejsca pracy oraz szczegółowego opisu odrzucano wszystkie oferty, które miały ten sam opis, najpierw w ramach jednego portalu (wykr. 2), a następnie dla grupy portali.

WYKR. 2. OFERTY PRACY POWTARZAJĄCE SIĘ W MAJU 2017 R.

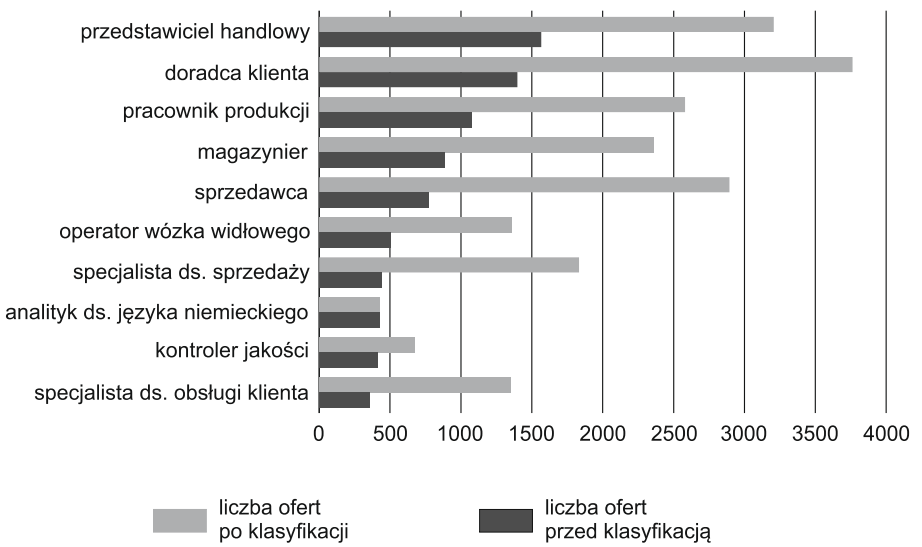


Ź r ó d ł o: opracowanie własne na podstawie wybranego portalu z ofertami pracy.

Innym wyzwaniem związanym z analizą ofert pracy jest dokonanie właściwej klasyfikacji danych. Właściciele portali internetowych zamieszczających ogłoszenia o pracy pozwalają pracodawcom na uwzględnienie dowolnych informacji w opisie ofert pracy. W związku z tym konieczne jest wykorzystanie metod text miningu, czyli metod eksploracji tekstu, w celu weryfikacji i przyporządkowania oferty pracy do właściwej kategorii według Klasyfikacji Zawodów i Specjalności (KZiS), bazującej na międzynarodowej klasyfikacji International Standard Classification of Occupation (ISCO). Metodą wspierającą właściwe odwzorowanie oferty pracy w klasyfikacji jest m.in. nadzorowane uczenie maszynowe, którego również nie można uznać za doskonałe narzędzie ze względu na dysproporcje w zbiorze treningowym w zakresie różnych zawodów. Przykładowo w zbiorze treningowym może być 100 obserwacji dla zawodu nauczyciel języka angielskiego i 5 tys. obserwacji dla zawodu programista. Oznacza to, że niektóre oferty pracy mogą mieć większą szansę na ich właściwe zaklasyfikowanie w stosunku do ofert rzadziej występujących w zbiorach treningowych. Skłania to bardziej do wykorzystywania metod deterministycznych, bazujących na wyrażeniach regularnych w text miningu, niż uczenia maszynowego. Wyrażenia regularne pozwalają na wyszukiwanie danych tekstowych według ustalonego wzorca, np. „angielsk*", co umożliwia wyeliminowanie problemów związanych z odmianą wyrazów w języku polskim.

Na wykr. 3 przedstawiono różnice pomiędzy liczbą ofert pracy przed zastosowaniem metod text miningu i po ich zastosowaniu. W celach testowych przygotowano również zbiór treningowy do uczenia maszynowego nadzorowanego w sposób ręczny, tj. przypisując nazwę oferty pracy do odpowiedniego kodu KZiS. Z przeprowadzonej analizy wynika, że nie jest możliwe uzyskanie precyzji klasyfikowania ofert prac na poziomie wyższym niż 90%.

WYKR. 3. EFEKT ZASTOSOWANIA TEXT MININGU DO KLASYFIKACJI OFERT PRACY



Ź r ó d ł o: opracowanie własne.

Jak zaprezentowano na wykr. 3, po zastosowaniu właściwego algorytmu klasyfikowania ofert pracy, z wykorzystaniem wcześniej wspomnianych metod text miningu i wyrażeń regularnych, można uzyskać inną liczbę ofert pracy w jednym zawodzie. Oznacza to, że niektóre oferty pracy przed zastosowaniem tych metod nie były przyporządkowane do zawodu zbliżonego do pozycji z KZiS.

Brak spójności dotyczy także wymiaru terytorialnego. Na portalach nie obowiązuje ujednolicona forma prezentacji, np. zgodna z rejestrem TERYT. Pojawiają się ogłoszenia, które nie zawierają określonej lokalizacji miejsca pracy (1–2% wszystkich ofert) lub jest ona bardzo ogólna, typu „cała Polska", „województwo pomorskie" czy „Gdańsk, Toruń, Warszawa".

Do identyfikacji zawodów w ofertach pracy wykorzystano wyrażenia regularne, które badały występowanie słów kluczowych z KZiS, np. „programista". Do-

datkowo utworzono rozwinięcie tego słownika o istotne słowa kluczowe, np. „Java" lub „Python" dla zawodu programisty. To niezbędny zabieg, gdyż w ofertach pracy nie zawsze występuje słowo „programista", a zamiast tego pojawia się np. „Java Developer". Przetwarzane dane oczyszczono wcześniej w procesie text miningu, tj. usunięto wyrazy niemające znaczenia dla analizy tekstu (ang. *stop words*). Ponadto sprowadzono słowa do postaci słownikowej, czyli do mianownika, następnie wyodrębniono rdzeń słów kluczowych dla danych zawodów, np. „programi*", który obejmuje takie formy, jak „programistów", „programista", „programiści" itd.

Najważniejsza kwestia w prowadzonym badaniu dotyczy porównywalności danych ze źródeł typu big data z danymi pozyskiwanymi w statystyce publicznej. Analiza poniższych danych ma jedynie charakter poglądowy, gdyż oferta pracy zamieszczona w internecie nie jest tożsama z wolnym miejscem pracy, które stanowi przedmiot badania GUS (tabl. 1).

TABL. 1. WOLNE MIEJSCA PRACY WEDŁUG SPRAWOZDANIA Z-05 A OFERTY PRACY W ŹRÓDŁACH BIG DATA W II KWARTALE 2017 R.

Wyszczególnienie	Z-05 (wakaty)	Big data (oferty pracy)
	w tys.	
O g ó ł e m[a]	**122,0**	**110,0**
w tym:		
Woj. mazowieckie	27,9	23,2
Pracownicy usług i sprzedawcy	14,5	15,3

a Dane za II kwartał 2018 r. wynoszą odpowiednio: 164,7 i 131,0.

Ź r ó d ł o: opracowanie własne oraz GUS (2018).

Należy zaznaczyć, że dane w kolumnie big data mają charakter eksperymentalny i wymagają nieustannej ingerencji w metody służące do przetwarzania i analizy danych. Jest to związane z dużą zmiennością źródeł danych big data, co zostanie wyjaśnione w kolejnej części artykułu. Warto jednak zaznaczyć, że po przetworzeniu danych typu big data, polegającym m.in. na usunięciu duplikatów i zadbaniu o spójność klasyfikacji, określono skalę brakujących ofert pracy – dane porównywano grupami zawodów. Za brakujące oferty pracy uznano te, które nie są publikowane w internecie. Liczby uzyskane przy wykorzystaniu narzędzi big data są dużo niższe niż w przypadku danych pozyskiwanych na podstawie sprawozdania Z-05. Jest to znane zjawisko w analizie źródeł internetowych big data, które występuje również w innych zastosowaniach web scrapingu, np. do analizy cen produktów. Zwykle ceny produktów dostępnych w internecie są niższe niż w sklepach stacjonarnych. Odwrotnie jest w przypadku cen na rynku nieruchomości, gdzie zwykle cena ofertowa przewyższa cenę transak-

cyjną. Zaskakuje jednak to, że liczba ofert pracy dla grupy „pracownicy usług i sprzedawcy" przewyższa liczbę wolnych miejsc pracy podawaną na podstawie danych ze sprawozdania Z-05. Może to mieć związek z bardzo dużą rotacją pracowników zatrudnionych na tych stanowiskach.

Ze względu na różnicę definicji wakatu w przypadku sprawozdania Z-05 oraz oferty pracy online w przypadku źródła big data uzyskane liczby nie mogą być bezpośrednio porównywane. Jednak stosowanie źródeł typu big data umożliwi śledzenie trendów w zakresie popytu na rynku pracy, dostarczając bieżącej informacji w bardzo szybkim czasie.

DANE SZCZEGÓŁOWE

Rozpatrując przedstawione powyżej wyniki, przede wszystkim należy wziąć pod uwagę wnioski dotyczące stabilności źródeł danych, tj. możliwości stałego pobierania danych ze stron internetowych o zbliżonym poziomie jakości. Wyniki badania pokazujące wahania w zakresie zmiany liczby ofert pracy rok do roku zaprezentowano w tabl. 2.

TABL. 2. UŚREDNIONA LICZBA OFERT PRACY W II i III KWARTALE

Portale		2017	2018
		w tys.	
Ogólnopolskie:	1	140	64
	2	124	19
	3	46	86
	4	39	54
Branżowe:	1	1,2	0,6
	2	0,4	0,8
Regionalny		9,9	8,5

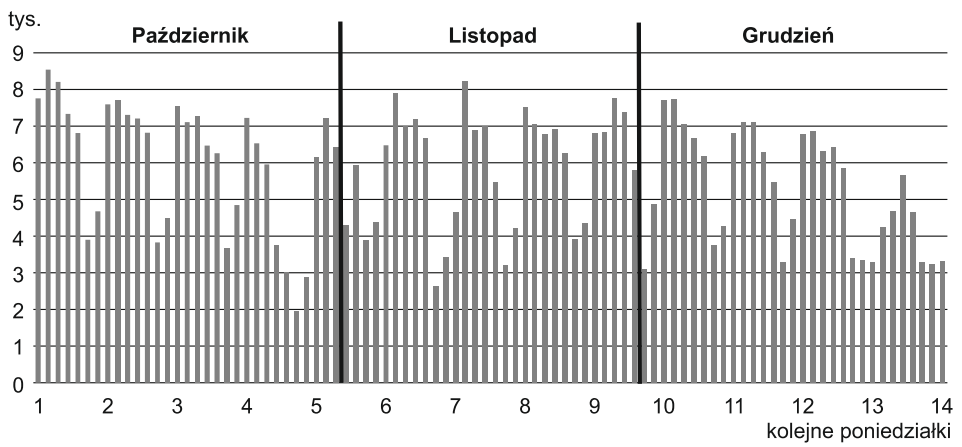
U w a g a. Cyfry w kolumnie Portale oznaczają kolejne badane portale.
Ź r ó d ł o: opracowanie własne.

Jak wynika z powyższej tabeli, zbiory danych wykazują dużą niestabilność – niektóre portale odnotowały niemal dwukrotny spadek liczby bieżących ofert pracy. Przyczyn należy upatrywać w odejściu wielu użytkowników tych portali do innych serwisów, oferujących więcej ofert. Ponadto coraz częściej praca jest oferowana poprzez media społecznościowe (np. Twitter, Facebook), bezpośrednio na stronie firmy czy na targach pracy dla profesjonalistów. Niektóre firmy zatrudniają headhunterów, którzy zajmują się poszukiwaniem odpowiednich osób poprzez portale społecznościowe, takie jak LinkedIn. W ramach tego portalu użytkownicy mogą zadeklarować swoje preferencje odnośnie do ewentualnej

zmiany pracy i otwartości na oferty headhunterów. Warto również podkreślić, że coraz więcej firm komercyjnych pobiera dane z internetu, a następnie udostępnia je w formie analiz. Istotnym przykładem wykorzystania ofert pracy online w celu budowy statystyk jest narzędzie do przeglądania ofert pracy online o nazwie Skills Online Vacancy Analysis Tool for Europe (Skills-OVATE) opracowane przez Europejskie Centrum Rozwoju Kształcenia Zawodowego (Cedefop)[5].

Brak stabilności portali internetowych, który nie pozwala uznać ich za niezmienne i stałe źródła danych, zdaje się dodatkowo potwierdzać analiza danych ujętych w układzie dziennym. Na wykr. 4 przeanalizowano dane z IV kwartału 2018 r. dotyczące jednego z większych portali internetowych zamieszczających oferty pracy.

WYKR. 4. NOWE OFERTY PRACY OPUBLIKOWANE W IV KWARTAŁ 2018 R.

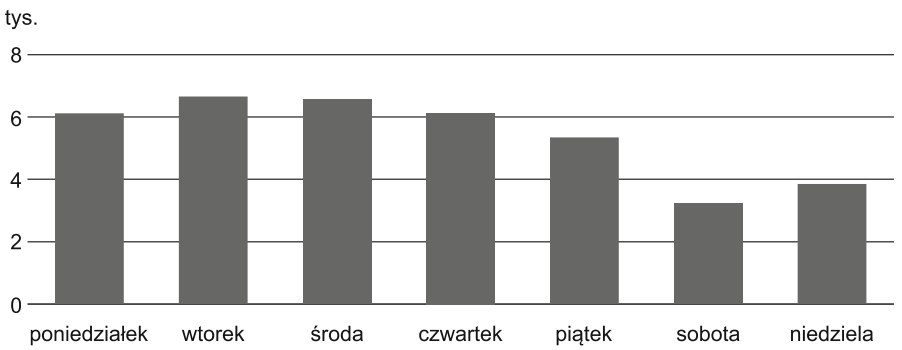


Ź r ó d ł o: opracowanie własne na podstawie danych z wybranego portalu z ofertami pracy.

Dane na wykr. 4 pokazują, że w zależności od dnia miesiąca liczba nowych ofert pracy może różnić się o ponad 50%. Dzienne zmiany dotyczące liczby ofert mogą sięgać kilkunastu procent. Oznacza to, że wyniki powinny być uśredniane w skali miesiąca/kwartału, aby w sposób wiarygodny zobrazować popyt na pracę. Istotną wartość dodaną stanowi analiza, kiedy pojawia się najwięcej ofert pracy w skali roku, kwartału czy nawet dnia tygodnia (wykr. 5), ale te informacje mogą być bardziej przydatne dla osób poszukujących pracy niż dla statystyki publicznej.

[5] https://www.cedefop.europa.eu (dostęp: 21.06.2019).

WYKR. 5. NOWE OFERTY PRACY WEDŁUG DNIA TYGODNIA – DANE UŚREDNIONE Z IV KWARTAŁU 2018 R.



Ź r ó d ł o: opracowanie własne na podstawie danych z wybranego portalu z ofertami pracy.

Po wyborze właściwego źródła danych zawarte w nim oferty są wstępnie przetwarzane. Podczas tego procesu następuje:

- deduplikacja danych, czyli usunięcie takich samych ofert pracy, powtarzanych wielokrotnie;
- ręczne odwzorowywanie KZiS lub wykorzystanie w tym celu uczenia maszynowego;
- dopasowanie miejscowości do województw zgodnie z rejestrem TERYT.

Trzeba pamiętać, że łączenie źródeł danych zwiększa liczbę duplikatów. Należy się zatem skoncentrować na możliwie pełnych zbiorach danych. Zagadnienie reprezentacyjności ma istotny wpływ na możliwość zastosowania metody web scrapingu w statystyce publicznej. Wykorzystanie narzędzi big data można rozpatrywać pod kątem wzbogacenia obecnie dostępnych danych statystycznych oraz uchwycenia trendów w możliwie najszybszym czasie, nawet kilka dni po zebraniu danych. W celu uniknięcia problemów z niestabilnością źródeł danych oraz precyzyjnego określenia metodologii wskazane jest nawiązanie współpracy z właścicielami serwisów internetowych.

PODSUMOWANIE

Na podstawie przeprowadzonych badań eksperymentalnych stwierdzono, że internetowe źródła danych charakteryzują się dużą niestabilnością i niepewną jakością zamieszczanych w nich informacji. Oferty pracy publikowane w internecie mogą być nieaktualne. Dodatkowo dowolność nazewnictwa zawodu w ofercie pracy sprawia, że konieczne jest ręczne odwzorowywanie zawodu w celu uzyskania zgodności z KZiS. Jedną z metod wspierających może być nadzorowane uczenie maszynowe. Niekiedy utrudnieniem jest brak jednoznacznego przyporządkowania terytorialnego, co również wymaga zastosowania wyrażeń

regularnych oraz metod text miningu w celu ekstrakcji takiej informacji. Co więcej, zastosowanie wielu źródeł danych może prowadzić do powstawania duplikatów, gdyż te same oferty mogą być powtarzane na wielu portalach.

Co istotne, wiele krajów europejskich pracuje nad rozwiązaniami pozwalającymi analizować oferty pracy online. Jako przykład mogą posłużyć prace prowadzone w ramach grantów ESSNet czy też informacje gromadzone przez Cedefop, gdzie oferty pracy są zbierane już od wielu lat. Zaletą pokazanego w niniejszym artykule rozwiązania jest bardzo szybkie dostarczanie danych wynikowych, które mogą zobrazować trendy w zakresie zmiany liczby ofert pracy. Wciąż jednak niepewność co do źródeł danych przyczynia się do braku jednoznacznej odpowiedzi, czy tego rodzaju dane mogą być wykorzystywane przez statystykę publiczną jako oficjalne dane statystyczne.

Zaprezentowane szacunki mają wyłącznie charakter eksperymentalny i nie powinny być traktowane jako oficjalne dane statystyczne ze względu na niestabilność źródeł danych i prawdopodobną konieczność weryfikacji w przyszłości założeń do algorytmów. W artykule zwrócono uwagę na możliwości stosowania źródeł big data do przetwarzania i analizowania danych statystycznych. Należy przy tym zaznaczyć, że oferty pracy zamieszczane w internecie nie są równoznaczne z wolnymi miejscami pracy, które przedsiębiorstwa wykazują w sprawozdawczości dla GUS. Tym samym istnieje konieczność zdefiniowania nowych terminów, pozwalających rozróżnić dwa odrębne pojęcia, jakimi są oferta pracy publikowana w internecie i wolne miejsce pracy.

BIBLIOGRAFIA

Beręsewicz, M., Szymkowiak, M. (2015). Big data w statystyce publicznej – nadzieje, osiągnięcia, wyzwania i zagrożenia. *Ekonometria*, *2*(48), 9–22. DOI: 10.15611/ekt.2015.2.01.

Braaksma, B., Zeelenberg, K. (2015). "Re-make/Re-model": Should big data change the modelling paradigm in official statistics? *Statistical Journal of the IAOS*, *31*(2), 193–202. DOI: 10.3233/sji-150892.

Daas, P. J. H., Puts, M. J., Buelens, B., van den Hurk, P. A. M. (2015). Big Data as a Source for Official Statistics. *Journal of Official Statistics*, *31*(2), 249–262. DOI: https://doi.org/10.1515/jos-2015-0016.

Douglas, L. (2001). *3D Data Management: Controling Data Volume, Velocity and Variety*. Pobrane z: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.

Gałecka-Burdziak, E., Pater, R. (2015). Ile jest wolnych miejsc pracy w Polsce? *Gospodarka Narodowa*, *279*(5), 171–186. DOI: https://doi.org/10.33119/GN/100855.

GUS. (2018). *Popyt na pracę w 2017 r*. Warszawa: Główny Urząd Statystyczny.

GUS. (2019). *Popyt na pracę w 2018 r*. Warszawa: Główny Urząd Statystyczny.

Hackl, P. (2016). Big Data: What can official statistics expect? *Statistical Journal of the IAOS*, *32*(1), 43–52. DOI: 10.3233/SJI-160965.

Kitchin, R. (2015). The opportunities, challenges and risks of big data for official statistics. *Statistical Journal of the IAOS*, *31*(3), 471–481. DOI: 10.3233/SJI-150906.

Kureková, L. M., Beblavý, M., Thum-Thysen, A. (2015). Using online vacancies and web surveys to analyse the labour market: a methodological inquiry. *IZA Journal of Labor Economics*, *4*(18), 1–20. DOI: 10.1186/s40172-015-0034-4.

Maślankowski, J. (2014). Data Quality Issues Concerning Statistical Data Gathering Supported by Big Data Technology. W: S. Kozielski, D. Mrozek, P. Kasprowski, B. Małysiak-Mrozek, D. Kostrzewa (red.). *Beyond Databases, Architectures and Structures* (s. 92–101) Cham: Springer.

Miller, S. (2014). Collaborative Approaches Needed to Close the Big Data Skills Gap. *Journal of Organization Design*, *3*(1), 26–30. DOI: 10.7146/jod.9823.

Rousidis, D., Garoufallou, E., Balatsoukas, P., Sicilia, M. (2014). Metadata for Big Data: a preliminary investigation of metadata quality issues in research data repositories. *Information Services &Use*, *34*(3–4), 279–286. DOI: 10.3233/ISU-140746.

Shahin, S. (2016). A Critical Axiology for Big Data Studies. *Palabra Clave*, *19*(4), 972–996. DOI: 10.5294/pacla.2016.19.4.2.

Vale, S. (2015). International collaboration to understand the relevance of Big Data for official statistics. *Statistical Journal of the IAOS*, *31*(2), 159–163. DOI: 10.3233/sji-150889.

Wiadomości Statystyczne. The Polish Statistician, 2019, vol. 64, 9, 75–77
In memoriam

Doktor Dariusz Parys (1963–2018)

Rok temu, 1 września 2018 r., zmarł w Łodzi dr Dariusz Parys, statystyk, badacz i dydaktyk, przez ponad 30 lat związany z Uniwersytetem Łódzkim (UŁ).

W 1982 r. podjął studia matematyczne na Wydziale Matematyki, Fizyki i Chemii UŁ. W 1987 r. zdobył tytuł magistra matematyki na podstawie pracy *Mocne prawo wielkich liczb Marcinkiewicza w algebrach von Neumana*. W 1988 r. podjął pracę na stanowisku asystenta w Instytucie Ekonometrii i Statystyki w Zakładzie Ekonometrii (od 1991 r. w Zakładzie Metod Statystycznych, a od 1992 r. w Katedrze Metod Statystycznych).

Fot. Archiwum autora

Szeroko rozwijał swoje zainteresowania teorią i zastosowaniem statystyki matematycznej w naukach ekonomiczno-społecznych oraz przyrodniczych i medycznych. Brał czynny udział w konferencjach, m.in. European Meeting of Statisticians w Barcelonie (1991), International Statistical Institute we Florencji (1993), Operation Research w Kolonii (1993), Minisymposium in Honour of Jaroslav Hajek na Uniwersytecie im. Karola w Pradze (1994) i Międzynarodowej Konferencji Biometrycznej w Poznaniu (1994). Jego wysoka aktywność była szczególnie zauważalna w organizacji międzynarodowej konferencji Wielowymiarowa Analiza Statystyczna w wielu jej edycjach począwszy od 1990 r. Pracował na stażach naukowych na University of Nijmegan w Holandii (1993) oraz University of Maryland (1996) i State University of Texas w Stanach Zjednoczonych (2003).

Zwieńczeniem osiągnięć naukowych z tego okresu była rozprawa doktorska *Testy wielokrotne i ich zastosowania*, przygotowana pod moim kierunkiem w 1997 r. Praca ta dotyczyła problemów wielowymiarowej analizy informacji statystycznych, a w szczególności metod wnioskowania statystycznego. Przedmiotem badań były porównania wielokrotne, wynikające z postawienia hipotezy o równości efektów zabiegów w analizie wariancji lub o zerowaniu się wektorów współczynnika regresji w klasycznym liniowym modelu regresji.

Dariusz Parys brał udział w grantach *Nieparametryczne metody wnioskowania statystycznego oparte na porównaniach wielokrotnych* oraz *Metody wnioskowania statystycznego oparte na porównaniach wielokrotnych* przyznanych

przez Komitet Badań Naukowych. Rezultatem drugiego z nich była wydana w 2007 r. monografia *Statystyczne metody wnioskowania wielokrotnego* (współautor: Czesław Domański), która zdobyła nagrodę zespołową Rektora Uniwersytetu Łódzkiego w 2008 r.

Po obronie pracy doktorskiej podjął badania stanowiące kontynuację dociekań zawartych w dysertacji. W przygotowywanej rozprawie habilitacyjnej *Kroczące metody porównań wielokrotnych i ich zastosowania w naukach społeczno-ekonomicznych* omawiał porównania wielokrotne między parametrami będącymi przedmiotem analizy wariancji i regresji. W procedurach porównań wielokrotnych istotną wadą jest brak kontroli nad popełnianym błędem wnioskowania. Procedury testowe kroczące zapewniają taką kontrolę, utrzymując błąd wnioskowania na ustalonym poziomie. W pracy dr. Parysa zawarte zostały liczne modyfikacje procedur kroczących o większej mocy niż istniejące.

Z powodu kłopotów zdrowotnych Dariusz Parys nie ukończył rozprawy habilitacyjnej. Niemniej w ostatnich latach, wraz z poprawą zdrowia, jego aktywność naukowa wzrosła, czego przejawem były wystąpienia na kilku międzynarodowych konferencjach.

Jako nauczyciel akademicki został dostrzeżony zarówno przez kolegów, jak i studentów UŁ. Był członkiem Rady Wydziału Ekonomiczno-Socjologicznego UŁ, a w latach 2009–2011 – członkiem Senatu UŁ. W 2010 r. w ramach programu Sokrates prowadził wykłady ze statystyki matematycznej na Uniwersytecie Santiago de Compostela w Hiszpanii. Był ceniony jako wykładowca rachunku prawdopodobieństwa i statystyki matematycznej w ramach studiów w Polsko-Amerykańskim Centrum Zarządzania funkcjonującym przy Wydziale Zarządzania UŁ. Prowadził także zajęcia z matematyki, matematyki finansowej, statystyki opisowej, planowania eksperymentów, teorii podejmowania decyzji, statystyki nieparametrycznej, statystycznej kontroli jakości, teorii gier, analizy matematycznej i algebry liniowej.

Jego wyjątkowy talent dydaktyczny docenili studenci, którzy w ramach konkursu zorganizowanego przez AIESEC nagrodzili go tzw. Nobelkiem w 2000 r. w kategorii „Wykładowca z Największym Poczuciem Humoru", a w 2001 r. w kategorii „Najlepszy Wykładowca na kierunku Finanse i Bankowość".

Należał do wielu towarzystw naukowych, m.in. Polskiego Towarzystwa Statystycznego (PTS), w którym pełnił funkcję wiceprzewodniczącego oddziału łódzkiego w latach 2010–2018 i członka Komisji Rewizyjnej, Polskiego Towarzystwa Biometrycznego, Polskiego Towarzystwa Ekonomicznego oraz Międzynarodowego Towarzystwa Bernoulli Society.

Jego dorobek naukowy obejmuje blisko 90 artykułów i streszczeń wystąpień z konferencji krajowych i zagranicznych.

Doktor Dariusz Parys został uhonorowany Złotą Odznaką Uniwersytetu Łódzkiego za wybitne zasługi w pracy naukowo-badawczej i dydaktycznej. Za dzia-

łalność w PTS został wyróżniony Medalem z okazji 90-lecia Polskiego Towarzystwa Statystycznego.

Jego życzliwość, talent i kultura osobista zjednały mu przyjaciół w wielu środowiskach w kraju i za granicą. Cieszył się zasłużonym szacunkiem i sympatią. Żył cicho i spokojnie. Tak też odszedł.

Czesław Domański (Uniwersytet Łódzki)

Wiadomości Statystyczne. The Polish Statistician, 2019, vol. 64, 9, 78–82
Informacje. Recenzje. Dyskusje / Information. Reviews. Discussions

Konferencja naukowa Metodologia Badań Statystycznych MET2019

Między 3 a 5 lipca br. odbyła się w Warszawie konferencja naukowa Metodologia Badań Statystycznych MET2019, zorganizowana przez Główny Urząd Statystyczny (GUS) we współpracy z Polskim Towarzystwem Statystycznym (PTS). Konferencja zgromadziła blisko 250 osób z kilku krajów.

W roli keynote speakerów wystąpiło dwóch wybitnych statystyków – prof. Graham Kalton z USA (Westat) i prof. Danny Pfeffermann z Izraela. Kalton zdobył międzynarodowe uznanie dzięki dorobkowi na polu metodologii badań sondażowych. Jest jednym z twórców Joint Program in Survey Methodology prowadzonego na University of Maryland, członkiem m.in. American Statistical Association, American Association for the Advancement of Science, International Statistical Institute oraz National Associate of the National Academy of Sciences (NAS). Pfeffermann specjalizuje się m.in. we wnioskowaniu analitycznym ze złożonych badań reprezentacyjnych, korektach sezonowych i szacowaniu trendów oraz w statystyce małych obszarów. Jest profesorem statystyki na Uniwersytecie w Southampton w Wielkiej Brytanii oraz emerytowanym profesorem Uniwersytetu Hebrajskiego w Jerozolimie, a także członkiem American Statistical Association i International Statistical Institute. Stoi na czele izraelskiego Centralnego Biura Statystycznego (Central Bureau of Statistics, CBS).

Program konferencji ukierunkowany był na problemy metodologiczne badań statystycznych w kontekście zmieniających się warunków funkcjonowania statystyki oficjalnej. Wystąpienia dotyczyły zarówno innowacyjnych przedsięwzięć statystyki publicznej, jak i sposobów komunikacji i udostępniania wyników badań, cechujących się wysoką jakością i wiarygodnością. Wydarzenie ma zainicjować cykl spotkań ekspertów statystyki publicznej z przedstawicielami środowisk naukowych, praktykami zarządzania i biznesu oraz użytkownikami danych statystycznych.

Tematyka konferencji obejmowała zagadnienia takie, jak:

- statystyka matematyczna, metoda reprezentacyjna i statystyka małych obszarów, statystyka ludności, społeczna, gospodarcza oraz statystyka regionalna;
- analiza i klasyfikacja danych, big data i dane statystyczne, statystyka polska na arenie międzynarodowej oraz miary i wskaźniki w statystyce publicznej;
- historia polskiej statystyki, metody statystyczne w badaniach historycznych, komunikacja statystyki i edukacja statystyczna.

Konferencję otworzył prezes GUS dr Dominik Rozkrut, który ogłosił Manifest 5 „O” na rzecz otwartości i transparentności w statystyce publicznej. Zaprezentował także nowy Portal Naukowy GUS, na którym zamieszczane są czasopisma i monografie naukowe wydawane przez Urząd. Powstanie również portal dotyczący programu European Master in Official Statistics (EMOS) jako platfor-

ma komunikacji, promocji i wymiany informacji o wydarzeniach związanych z EMOS. Natomiast na stronie dekompozycje.stat.gov.pl będzie można projektować i monitorować strategie rozwoju regionalnego, a także analizować wzrost i różnice w poziomie rozwoju gospodarczego województw i makroregionów.

W pierwszym dniu konferencji odbyło się dziewięć sesji, podczas których wygłoszono 31 referatów. Wystąpienie prof. dr hab. Grażyny Trzpiot i prof. dr. hab. inż. Jacka Szołtyska dotyczyło miast przyjaznych seniorom. Autorzy przyjęli za punkt wyjścia definicję WHO, wskazującą cztery główne kryteria, według których można ocenić, czy dane miasto jest przyjazne osobom starszym, są to: zasady, usługi, otoczenie i struktury, które wspierają proces aktywnego starzenia się. W ramach badania stworzono model opisowy, a następnie dokonano oceny wybranych miast Polski, wykorzystując w tym celu podejście taksonomiczne.

Prof. dr hab. Czesław Domański i dr hab. Alina Jędrzejczak poruszyli temat problemów etycznych w procesie generowania danych w internecie. Zwrócili uwagę na towarzyszące badaniom statystycznym przekonanie, że dostęp do wiarygodnych informacji niesie społeczne korzyści. Statystycy powinni brać pod uwagę prawdopodobieństwo konsekwencji zbierania i upowszechniania danych, eliminując możliwe do przewidzenia przypadki ich nieprawidłowej interpretacji lub niewłaściwego wykorzystania.

Zagadnienie cząstkowych indeksów polaryzacyjnych omówili dr Jan Zwierzchowski i prof. dr hab. Tomasz Panek. W okresie objętym przeprowadzonym przez nich badaniem gospodarstwa domowe znajdujące się w środkowej klasie dochodowej przesuwały się w kierunku niższej lub wyższej klasy dochodowej, natomiast gospodarstwa domowe znajdujące się w klasach niższej i wyższej przemieszczały się w kierunku mediany rozkładu dochodów. Co więcej, konwergencja obserwowana w niższych i wyższych klasach dochodowych była silniejsza niż polaryzacja w środkowej klasie dochodowej. Przepływy gospodarstw domowych do środkowej klasy dochodowej były wyższe niż wypływy z tej klasy.

Referat *Wkład polskiej myśli statystycznej do statystyki światowej* wygłosili Czesław Domański i prof. dr hab. Włodzimierz Okrasa. Przedstawili dorobek wybranych statystyków z Polski, których wkład w rozwój statystyki wart jest uwidocznienia i popularyzacji.

Referat dr. Marka Cierpiał-Wolana i Sebastiana Wójcika dotyczył budowy innowacyjnego systemu integracji danych w dziedzinie turystyki. Pojawienie się źródeł danych takich jak big data wymaga nowego podejścia do tworzenia wiarygodnych informacji statystycznych. Ma to związek z rozproszonym charakterem tych źródeł, a także z zamiarem wykorzystania ich na dużą skalę w oficjalnych statystykach. To podejście, często określane terminem *trusted smart statistics*, będzie w najbliższych latach stanowiło wyzwanie dla oficjalnej statystyki. Autorzy podkreślili, że innowacyjne metody zbierania i łączenia danych pozwalają na zwiększenie poziomu ich dezagregacji, poprawę jakości operatów oraz

opracowywanie szybkich szacunków w dziedzinie turystyki, co odegra kluczową rolę w rozwoju spójnego systemu informacji turystycznej.

W drugim dniu konferencji odbyło się 12 sesji ukierunkowanych na zagadnienia związane m.in. z tematyką statystyki gospodarczej, big data i danych statystycznych oraz statystyki ludności. Wygłoszono 47 referatów. Dr Katarzyna Brzozowska-Rup i dr Mariola Chrzanowska omówiły badania zjawiska szarej strefy. Zaprezentowały koncepcję oraz zastosowanie modeli strukturalnych zmiennych ukrytych (Structural Equation Models, SEM). Mogą one służyć do bezpośredniego pomiaru aktywności gospodarczych prowadzonych bez pełnej wiedzy państwa, co w konsekwencji skłania do wykorzystania modeli ze zmiennymi nieobserwowalnymi i niemierzalnymi. Prelegentki podkreśliły istnienie niewykorzystanego potencjału eksplikacyjnego modeli równań strukturalnych, w związku z czym należałoby podjąć próbę specyfikacji i rozszerzenia tych modeli w kontekście badań szarej strefy.

Wybrane problemy reprezentacyjnych badań powtarzalnych przedstawiły dr hab. Barbara Kowalczyk i dr Dorota Juszczak. Dokonały one przeglądu metod estymacji na podstawie danych otrzymanych z badań rotacyjnych. Następnie zaprezentowały wyniki teoretyczne oraz omówiły wyniki badań symulacyjnych.

Zagadnienia dotyczące wykorzystania bootstrapowanych modeli ekonometrycznych dla danych indywidualnych w prognozowaniu dotyczącym nowych lokalizacji w zbiorach big data poruszyła dr hab. Katarzyna Kopczewska. Typowe przestrzenne modele ekonometryczne, takie jak SDM, SDEM czy SAC, szacowane dla danych punktowych geo-lokalizowanych cechują się ograniczoną możliwością prognozowania w przypadku nowych punktów w przestrzeni spoza próby. Przedstawiona nowatorska metodologia kalibruje zarówno przestrzeń, jak i model przy użyciu bootstrapu i teselacji. Bootstrap umożliwia kalibrację modelu ekonometrycznego bez potrzeby szacowania na całym zestawie danych. To rozwiązanie przetestowano na podstawie lokalizacji firm z rejestru REGON.

Dr hab. Mirosława Kaczmarek podjęła temat wykluczenia pokolenia Baby Boomers (BB) z rynku e-commerce. Prelegentka skupiła się na określeniu pozycji Polski na tle krajów Unii Europejskiej (UE) pod względem korzystania z handlu i usług dostępnych online przez osoby z pokolenia BB; zwróciła uwagę na płeć i wykształcenie jako czynniki determinujące wykluczenie z rynku e-commerce. Polska znajduje się w grupie 13 krajów, które charakteryzują się niskimi wskaźnikami korzystania z e-handlu, a na wykluczenie z rynku e-commerce osób z generacji BB istotnie wpływa właśnie poziom ich wykształcenia.

Trzeciego dnia konferencji odbyło się sześć sesji dotyczących m.in. statystyki regionalnej, statystyki gospodarczej i statystyki matematycznej, podczas których wygłoszono 26 referatów. Wystąpienie dr Katarzyny Cheby i dr hab. Iwony Bąk *Europejski Ranking Innowacyjności i jego regionalny wymiar – analiza taksonomiczna* dotyczyło problemu znacznego wewnętrznego zróżnicowania poziomu

innowacyjności krajów i regionów UE. Zauważono, że uśrednianie wyników uzyskiwanych w ramach poszczególnych badanych obszarów i przedstawianie ich w postaci jednego miernika w publikowanych przez Komisję Europejską raportach European Innovation Scoreboard i Regional Innovation Scoreboard może prowadzić do zbyt dużych uogólnień i opacznej klasyfikacji regionów do grup typologicznych.

Dominika Rogalińska i dr Marek Pieniążek wygłosili referat *Aplikacyjne aspekty kształtowania strumieni danych statystycznych w państwie na przykładzie badań przestrzennych*. Jak stwierdzili, obecny system przepływu danych charakteryzuje wysoki poziom złożoności, a dane gromadzone są według różnych metod i w różnych formach zapisu, a to ogranicza ich użyteczność w kompleksowym monitorowaniu. W związku z tym warto rozważyć wdrożenie rozwiązań stosowanych m.in. w krajach skandynawskich, w których statystyka publiczna pełni funkcję kluczowego węzła wymiany danych, dzięki czemu nakłady na realizację obowiązku sprawozdawczego zmniejszyły się, a jakość i spójność danych wzrosła. Przykładem działań podjętych w tym kierunku przez GUS są prace związane z budową Systemu Monitorowania Usług Publicznych (SMUP), który ma być pierwszym kompleksowym rozwiązaniem wspierającym doskonalenie świadczenia usług dla każdego obywatela w jednostkach samorządu terytorialnego. Prelegenci podkreślili, że usprawnienie systemu przekazywania danych jest szczególnie istotne na poziomie regionalnym, ponieważ dostęp do aktualnych i spójnych danych oraz możliwość ich przetwarzania stały się kluczowym czynnikiem rozwoju.

Dr Bohdan Wyżnikiewicz w referacie *Cena prawdy w statystyce* przybliżył zagadnienia związane z przypadkami podważania prawdziwości podawanych informacji statystycznych w polskiej statystyce publicznej. Zaproponował także schemat postępowania w odpowiedzi na krytykę danych statystycznych ze strony mediów i niektórych ekspertów.

Na dzień przed rozpoczęciem konferencji odbyły się warsztaty towarzyszące, które poprowadził Graham Kalton. Ich program obejmował omówienie metod próbkowania populacji stosowanych w ankietach, od prostego próbkowania losowego, poprzez systematyczne próbkowanie, grupowanie i próbkowanie wieloetapowe, po projekty paneli czy stosowanie powtarzających się ankiet. Kalton wiele uwagi poświęcił kwestiom związanym z praktycznym prowadzeniem badań sondażowych, w tym problemom wynikającym z braku odpowiedzi, niedokładności pomiarów lub pojawienia się duplikatów. Szczególnie owocna była dyskusja na temat prowadzenia badań statystycznych w krajach rozwijających się (np. afrykańskich), w których występują trudności związane z ankietowaniem zarówno lokalnej ludności, jak i miejscowej administracji. Jednym ze sposobów pokonania tego rodzaju problemów może być zastosowanie siatki selekcji, zaproponowanej po raz pierwszy przez Lesliego Kisha. Inne stosowane kroki to np. metoda Troldahl-Cartera czy metoda zaproponowana przez Louisa Rizzo. Kalton

omówił również metodę ACE (Address Coverage Enhancement), stworzoną w celu uzupełnienia metody ABS, oraz najnowsze trendy w badaniach sondażowych, obejmujące głównie metody web scrapingu stron takich jak Facebook czy Twitter.

Daniel Koprowicz (Urząd Statystyczny w Rzeszowie)

Wiadomości Statystyczne. The Polish Statistician, 2019, vol. 64, 9, 83—84
Informacje. Recenzje. Dyskusje / Information. Reviews. Discussions

Wydawnictwa GUS. Sierpień 2019

W sierpniowej ofercie wydawniczej warto zwrócić uwagę na raport ***Społeczna odpowiedzialność statystyki publicznej*** oraz zeszyt metodologiczny ***Współdziałanie jednostek samorządu terytorialnego z mieszkańcami***.

Społeczna odpowiedzialność statystyki publicznej to nowość wydawnicza prezentująca wybrane działania statystyki publicznej powiązane z ideą społecznej odpowiedzialności podjęte w 2018 r.

Raport został podzielony na sekcje tematyczne związane z ładem organizacyjnym, prawami człowieka, praktykami z zakresu prawa pracy, zagadnieniami środowiskowymi i konsumenckimi, a także zaangażowaniem statystyki w działania społeczne i rozwojem społeczności lokalnej.

Opracowanie wydano po polsku i dostępne jest na stronie internetowej GUS.

Zeszyt metodologiczny badania ***Współdziałanie jednostek samorządu terytorialnego z mieszkańcami*** dostarcza danych na temat zakresu i form zaangażowania obywateli i ich organizacji w procesy współzarządzania lokalną sferą publiczną w gminach. Włączenie tego badania do programu badań statystycznych, jako elementu sprawozdawczości z zakresu samorządu terytorialnego, to odpowiedź na rosnące zapotrzebowanie na informacje o zaangażowaniu obywateli w sprawy publiczne na poziomie wspólnoty lokalnej, związane ze zwiększającą się samoświadomością obywatelską i zmianami w ustawodawstwie.

Opracowanie, bazujące na pracy badawczej *Pozyskanie nowych wskaźników dotyczących realizacji usług publicznych z zakresu partycypacji społecznej*, zawiera opis zestawu danych dotyczących współdziałania jednostek samorządu terytorialnego z mieszkańcami, w tym określenie zakresu podmiotowego i przedmiotowego badania, wyszczególnienie zmiennych w badaniu, szczegółowe przedstawienie metody ich pozyskiwania oraz omówienie form prezentacji danych wynikowych.

Publikacja jest dostępna na stronie Urzędu w polskiej wersji językowej.

W sierpniu br. ukazały się ponadto:

- „Biuletyn statystyczny" nr 6/2019;
- *Ceny robót budowlano-montażowych i obiektów budowlanych (czerwiec 2019 r.)*;
- *Koniunktura w przetwórstwie przemysłowym, budownictwie, handlu i usługach 2000–2019 (sierpień 2019)*;
- *Produkcja ważniejszych wyrobów przemysłowych w lipcu 2019 r.*;
- *Regiony Polski 2019*;
- *Sytuacja demograficzna Polski do 2018 roku. Tworzenie i rozpad rodzin*;
- *Sytuacja społeczno-gospodarcza kraju w lipcu 2019 r.*;
- „Wiadomości Statystyczne" nr 8/2019 (699).

Wersje elektroniczne wszystkich publikacji GUS dostępne są na stronie https://stat.gov.pl/publikacje/publikacje-a-z/.

Justyna Gustyn (Główny Urząd Statystyczny)

Dla autorów
For the authors

(for information go to: https://ws.stat.gov.pl/ForAuthors)

W „Wiadomościach Statystycznych. The Polish Statistician" („WS") zamieszczane są artykuły o charakterze naukowym poświęcone teorii i praktyce statystycznej, które prezentują wyniki oryginalnych badań teoretycznych lub analitycznych wykorzystujących metody statystyki matematycznej, opisowej bądź ekonometrii. Ukazują się również artykuły przeglądowe, recenzje publikacji naukowych oraz inne opracowania informacyjne. W czasopiśmie publikowane są prace w języku polskim i angielskim.

Od 2007 r. „WS" znajdują się na liście polskich punktowanych czasopism naukowych Ministerstwa Nauki i Szkolnictwa Wyższego. Zgodnie z komunikatem MNiSW z dnia 31 lipca 2019 r. w sprawie wykazu czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych wraz z przypisaną liczbą punktów „WS" otrzymały 20 punktów.

„Wiadomości Statystyczne. The Polish Statistician" są udostępniane w następujących bazach indeksacyjnych i repozytoriach: POL-index, CEJSH, BazEkon oraz AGRO.

Za publikację artykułów na łamach „WS" autorzy nie otrzymują honorariów ani nie wnoszą opłat.

Zgłaszanie artykułów

Prace należy przesyłać drogą elektroniczną na adres: **redakcja.ws@stat.gov.pl**.

Artykuł powinien być utrzymany w formie bezosobowej i zawierać streszczenie, słowa kluczowe oraz kod/kody JEL. Tytuł, streszczenie i słowa kluczowe powinny być podane w języku polskim i angielskim. Jeżeli w pracy występują tablice, wykresy lub mapy, powinny być umieszczone w treści artykułu. W osobnym pliku należy podać dane do wykresów. **Prosimy o niestosowanie stylów i ograniczenie formatowania do wymogów redakcyjnych.** Więcej informacji – w podrozdziale *Wymogi redakcyjne* i następnych podrozdziałach.

Razem z artykułem należy przesłać skan oświadczenia (do pobrania ze strony internetowej czasopisma) o oryginalności pracy i niezłożeniu jej w innym wydawnictwie, zawierającego zgodę na przeniesienie autorskich praw majątkowych, numer ORCID oraz dane kontaktowe autora i afiliację zgłaszanego artykułu wraz ze wskazaniem proponowanego działu czasopisma. Oryginał oświadczenia należy wysłać na adres: Redakcja „Wiadomości Statystycznych. The Polish Statistician", Główny Urząd Statystyczny, al. Niepodległości 208, 00-925 Warszawa.

Załączenie skanu oświadczenia jest warunkiem poddania pracy ocenie wstępnej i skierowania do recenzji.

Przebieg prac redakcyjnych

Redakcja rozpoczyna postępowanie kwalifikujące artykuł do opublikowania po przedłożeniu przez autora oświadczenia o przeniesieniu praw majątkowych do artykułu.

Zgłoszony artykuł jest oceniany i opracowywany zgodnie ze schematem:

1. **Ocena wstępna**, dokonywana przez redakcję. Polega na weryfikacji naukowego charakteru artykułu oraz jego struktury i zawartości pod kątem wymogów redakcyjnych, a także zgodności tematyki z profilem czasopisma. Autor uzupełnia i poprawia artykuł stosownie do uwag redakcji, a w przypadku nieuwzględnienia danej uwagi uzasadnia swoje stanowisko. **Razem z poprawionym artykułem autor przesyła w osobnym pliku zanonimizowaną wersję artykułu, która jest kierowana do recenzji.** Anonimizacja polega na utajnieniu nazwiska autora (także we właściwościach pliku), usunięciu podziękowań i informacji o źródłach finansowania, a także innych informacji wskazujących na afiliację lub umożliwiających zidentyfikowanie autora.
2. **Ocena recenzentów**, dokonywana przez specjalistów w danej dziedzinie. Artykuł oceniają dwaj recenzenci spoza jednostki naukowej, do której afiliowana jest zgłoszona praca; w przypadku artykułu w języku angielskim co najmniej jeden recenzent jest afiliowany przy jednostce zagranicznej. W razie sprzecznych opinii dwóch recenzentów powoływany jest trzeci recenzent. Recenzenci kierują się kryteriami oryginalności i jakości opracowania zarówno w odniesieniu do treści, jak i formy.

   Autorzy artykułów, które otrzymały pozytywne oceny, wprowadzają poprawki zalecane przez recenzentów i dostarczają redakcji zmodyfikowaną wersję pracy. Jeśli pojawi się różnica zdań dotycząca zasadności proponowanych zmian, autorzy są zobligowani do uzasadnienia swojego stanowiska.
3. **Ocena dopuszczająca do publikacji**, dokonywana przez Kolegium Redakcyjne (KR) na podstawie recenzji, z uwzględnieniem opinii redaktorów tematycznego i merytorycznego. Polega m.in. na weryfikacji dokonania przez autora zmian w artykule stosownie do uwag recenzentów. Kolegium Redakcyjne ocenia artykuł pod względem poprawności i spójności merytorycznej oraz zaleca autorowi wprowadzenie poprawek, jeśli są one konieczne, aby praca spełniała wymogi czasopisma. W przypadku podjęcia decyzji o niepublikowaniu artykułu autorowi przysługuje prawo do odwołania. W tym celu powinien on skontaktować się z redakcją „WS" i przedstawić uzasadnienie. Ostateczna decyzja w tej sprawie należy do redaktora naczelnego.

   W „WS" publikowane są wyłącznie te artykuły, które otrzymają pozytywną ocenę na każdym z wymienionych etapów i zostaną poprawione przez autora zgodnie z otrzymanymi uwagami. W przypadku nieuwzględnienia danej uwagi autor jest zobligowany do uzasadnienia swojego stanowiska.
4. **Opracowanie redakcyjne, autoryzacja i korekta**. Artykuł zakwalifikowany do druku jest poddawany opracowaniu merytorycznemu i językowemu. Re-

dakcja zastrzega sobie prawo do zmiany tytułu i śródtytułów, modyfikowania tablic, wykresów i innych elementów graficznych oraz przeredagowania treści bez naruszenia zasadniczej myśli autora.

Po opracowaniu redakcyjnym artykuł jest przesyłany do autoryzacji. Tekst zatwierdzony przez autora, po składzie i łamaniu, jest poddawany korekcie i rewizji (II korekcie). Autor dokonuje korekty autorskiej tekstu na etapie rewizji. Wykresy i inne materiały graficzne są opracowywane na podstawie danych przekazanych przez autora i poddawane korekcie i rewizji. Autor dokonuje ich akceptacji na etapie rewizji.

W przypadku odkrycia błędów w opublikowanym artykule zamieszcza się na łamach „WS" sprostowanie, a artykuł w wersji elektronicznej jest poprawiany i umieszczany na stronie internetowej „WS" ze stosownym wyjaśnieniem.

Zasady etyki publikacyjnej COPE

Redakcja „WS" dokłada wszelkich starań, aby utrzymać najwyższe standardy etyczne, zgodnie z wytycznymi Komitetu ds. Etyki Publikacyjnej (COPE), dostępnymi na stronie internetowej www.publicationethics.org, oraz wykorzystuje wszystkie możliwe środki mające na celu zapobieżenie nadużyciom i nierzetelności autorskiej. Przyjęte zasady postępowania obowiązują autorów, redakcję, recenzentów i wydawcę.

Odpowiedzialność autorów

1. Artykuły naukowe kierowane do opublikowania w „WS" powinny zawierać precyzyjny opis badanych zjawisk i stosowanych metod oraz autorskie wnioski i sugestie dotyczące rozwoju badań i analiz statystycznych. Autorzy powinni wyraźnie określić cel artykułu oraz jasno przedstawić wyniki przeprowadzonej analizy. Prezentacja efektów badań statystycznych zaprojektowanych i przeprowadzonych przez autorów wymaga opisania zastosowanej w nich metodologii. W przypadku nowatorskich metod analizy pożądane jest podanie przykładu ilustrującego ich zastosowanie w praktyce statystycznej. Autorzy ponoszą odpowiedzialność za treści prezentowane w artykułach. W razie zgłaszania przez czytelników zastrzeżeń odnoszących się do tych treści autorzy są zobligowani do udzielenia odpowiedzi za pośrednictwem redakcji.
2. Na autorach spoczywa obowiązek zapewnienia pełnej oryginalności przedłożonych prac. Redakcja nie toleruje przejawów nierzetelności naukowej autorów, takich jak:
   - duplikowanie publikacji – ponowne publikowanie własnego utworu lub jego części;
   - plagiat – przywłaszczenie cudzego utworu lub jego fragmentu bez podania informacji o źródle;
   - fabrykowanie danych – oparcie pracy naukowej na nieprawdziwych wynikach badań;

- autorstwo widmo (*ghost authorship*) – nieujawnianie współautorów, mimo że wnieśli oni istotny wkład w powstanie artykułu;
- autorstwo gościnne (*guest authorship*) – podawanie jako współautorów osób o znikomym udziale lub niebiorących udziału w opracowaniu artykułu;
- autorstwo grzecznościowe (*gift authorship*) – podawanie jako współautorów osób, których wkład jest oparty jedynie na słabym powiązaniu z badaniem.

Autorzy deklarują w stosownym oświadczeniu, że zgłaszany artykuł nie narusza praw autorskich osób trzecich, nie był dotychczas publikowany i nie został złożony w innym wydawnictwie oraz że jest ich oryginalnym dziełem, i określają swój wkład w opracowanie artykułu. Jeżeli doszło do zaprezentowania podobnych materiałów podczas konferencji lub sympozjum naukowego, to podczas składania tekstu do publikacji w „WS" autorzy są zobowiązani poinformować o tym redakcję.

3. Autorzy są zobowiązani do podania w treści artykułu wszelkich źródeł finansowania badań będących podstawą publikacji.
4. Główną odpowiedzialność za rzetelność przekazanych informacji, łącznie z informacją na temat wkładu poszczególnych współautorów w powstanie artykułu, ponosi zgłaszający artykuł.
5. Autorzy zgłaszający artykuły do publikacji w „WS" biorą udział w procesie recenzji double-blind peer review, dokonywanej przez co najmniej dwóch niezależnych ekspertów z danej dziedziny. Po otrzymaniu pozytywnych recenzji autorzy wprowadzają zalecane przez recenzentów poprawki i dostarczają redakcji zaktualizowaną wersję opracowania wraz z poświadczeniem na piśmie uwzględnienia poprawek. Jeśli pojawi się różnica zdań co do zasadności proponowanych zmian, należy wyjaśnić, które poprawki zostały uwzględnione, a w przypadku ich nieuwzględnienia – uzasadnić swoje stanowisko.
6. Jeżeli autorzy odkryją w swoim maszynopisie lub tekście już opublikowanym błędy, nieścisłości bądź niewłaściwe dane, powinni o tym niezwłocznie poinformować redakcję w celu dokonania korekty, wycofania tekstu lub zamieszczenia odpowiedniego sprostowania. W przypadku korekty artykułu już opublikowanego jego nowa wersja jest zamieszczana na stronie internetowej „WS" wraz ze stosownym wyjaśnieniem.

Odpowiedzialność redakcji

1. Redakcja „WS" odpowiada za zorganizowanie i sprawny przebieg procesu wydawniczego, na który składają się: wstępna ocena zgłoszonego maszynopisu, ocena recenzentów (w przypadku artykułów naukowych), ocena KR, redakcja językowa, redakcja techniczna, skład i łamanie oraz korekta.
2. Redakcja ustala zasady obowiązujące w procesie wydawniczym, informuje jego uczestników o konieczności ich przestrzegania i egzekwuje je na każdym z jego etapów oraz dba o stałą aktualizację informacji na temat przyjętych zasad na stronie internetowej i na łamach czasopisma.
3. Redakcja nie może pozostawać w jakimkolwiek konflikcie interesów w odniesieniu do przyjmowanych artykułów. Przez konflikt interesów należy rozumieć

sytuację, w której wszelkie interesy lub związki (służbowe, finansowe lub inne) mogą mieć wpływ na obiektywną ocenę zgłoszonego maszynopisu lub decyzję o jego publikacji.

4. W celu przeciwdziałania nierzetelności naukowej redakcja wymaga od autorów złożenia oświadczenia, w którym deklarują oni, że zgłaszany artykuł nie narusza praw autorskich osób trzecich, nie był dotychczas publikowany i że jest ich oryginalnym dziełem, oraz określają swój wkład w opracowanie artykułu.
5. Podczas oceny wstępnej zgłoszony maszynopis jest weryfikowany przez redaktorów pod względem zgodności z celem i zakresem tematycznym czasopisma oraz spełniania wymogów redakcyjnych „WS", a także ewentualnych przejawów nierzetelności naukowej i możliwości wystąpienia konfliktu interesów.
6. Po ocenie wstępnej opracowania mające charakter naukowy przekazywane są do recenzji specjalistom z poszczególnych dziedzin. Redakcja jest odpowiedzialna za ustalenie spójnych kryteriów oceny artykułu oraz wymaga od recenzentów podpisania oświadczenia o przestrzeganiu zasad etyki recenzowania COPE (https://publicationethics.org/resources/guidelines-new/cope-ethical-guidelines-peer-reviewers) i niewystępowaniu konfliktu interesów. Informacje dotyczące maszynopisu mogą być przekazywane przez redakcję wyłącznie autorom, recenzentom, wydawcy lub innym doradcom redakcyjnym.
7. W przypadku podejrzenia nadużyć redakcja postępuje zgodnie z procedurami COPE.
8. Redakcja zapewnia, że zmiany dokonane w tekście na etapie prac redakcyjnych nie naruszają zasadniczej myśli autorów.
9. Kolegium Redakcyjne, podejmując decyzję o publikacji artykułu, kieruje się wyłącznie wynikiem dyskusji dotyczącej zgłoszonego artykułu, w której uwzględniane są oceny recenzentów oraz opinie redaktorów tematycznych i merytorycznych. Rezultat ten zależy od merytorycznej oceny wartości artykułu, jego oryginalności i jasności przekazu, a także od ścisłego związku z celem i zakresem tematycznym miesięcznika.
10. W przypadku podjęcia decyzji o niepublikowaniu przesłanego materiału redakcja nie może go w żaden sposób wykorzystać bez pisemnej zgody autora.

Odpowiedzialność recenzentów

1. Recenzenci przyjmują artykuł do recenzji tylko wtedy, gdy uznają, że:
   - posiadają odpowiednią wiedzę w określonej dziedzinie, aby rzetelnie ocenić pracę;
   - nie istnieje konflikt interesów w odniesieniu do autorów, przedstawionych w artykule badań i instytucji je finansujących, co potwierdzają w oświadczeniu;
   - mogą wywiązać się z terminu ustalonego przez redakcję, aby nie opóźniać publikacji.

2. Recenzenci są zobligowani do zachowania obiektywności i poufności oraz powstrzymania się od osobistej krytyki. Zawsze powinni uzasadnić swoją ocenę, przedstawiając stosowną argumentację.
3. W uzasadnionych przypadkach recenzenci powinni wskazać ważne dla wyników badań opublikowane prace, które w ich ocenie powinny zostać przywołane w ocenianym artykule.
4. W razie stwierdzenia wysokiego poziomu zbieżności treści recenzowanej pracy z innymi opublikowanymi materiałami lub podejrzenia innych przejawów nierzetelności naukowej recenzenci są zobowiązani poinformować o tym redakcję.
5. Po ukończeniu recenzji przechowywanie przesłanych przez redakcję materiałów (w jakiejkolwiek formie) oraz posługiwanie się nimi przez recenzentów jest niedozwolone.

Odpowiedzialność wydawcy

1. Materiały opublikowane w „WS" są chronione prawem autorskim.
2. Wydawca udostępnia pełną treść wszystkich artykułów w Internecie na zasadach otwartego dostępu, tj. bezpłatnie i bez technicznych ograniczeń. Użytkownicy mogą czytać, pobierać, kopiować, drukować i wykorzystywać do innych celów artykuły zamieszczone online, zgodnie z właściwymi przepisami o dozwolonym użytku, pod warunkiem wskazania źródła pochodzenia artykułu. Inne sposoby wykorzystania treści artykułów „WS" wymagają zgody wydawcy.
3. Wydawca deklaruje gotowość do opublikowania poprawek, wyjaśnień oraz przeprosin.

Wymogi redakcyjne

Zgodnie z wymogami czasopisma omawiany w artykule problem badawczy powinien być jednoznacznie zdefiniowany oraz istotny dla oceny zjawisk społecznych lub gospodarczych. Artykuł powinien zawierać wyraźnie określony cel badań, precyzyjny opis badanych zjawisk i stosowanych metod, uzyskane wyniki przeprowadzonej analizy oraz autorskie wnioski.

Zachęcamy do przygotowania pracy z wykorzystaniem szablonu artykułu „WS" do pobrania ze strony: https://ws.stat.gov.pl/ForAuthors.

Struktura i zawartość artykułu

Wymagane elementy artykułu:

1. Tytuł, autor.
2. Streszczenie (objętość do 1200 znaków ze spacjami, forma bezosobowa). W przypadku artykułu opisującego badanie empiryczne powinno zawierać: cel badania, przedmiot, okres i metodę badania, źródła danych, najważniejsze wnioski z badania. W przypadku artykułów o innym charakterze należy podać co najmniej cel artykułu, przedmiot i najważniejsze wnioski.

Streszczenie to podstawowe źródło informacji o artykule, warunkujące też decyzję czytelnika o zapoznaniu się z całą pracą. Dlatego powinno być przygotowane ze szczególną starannością i dbałością o umieszczenie w nim wszystkich wymaganych elementów.

3. Słowa kluczowe – najistotniejsze użyte w pracy pojęcia lub wyrażenia (nie mniej niż trzy). Słowa kluczowe powinny być zawarte w streszczeniu i/lub tytule.
4. Tłumaczenie tytułu, streszczenia i słów kluczowych (na język angielski w przypadku artykułu napisanego w języku polskim, a na język polski w przypadku artykułu napisanego w języku angielskim).
5. Kod/kody z klasyfikacji Journal of Economic Literature (JEL).
6. W przypadku artykułu opisującego badanie empiryczne wymagane są następujące części:
   - wprowadzenie, zawierające: cel badania, uzasadnienie podjętego problemu badawczego i odniesienie do literatury przedmiotu, chyba że przegląd literatury stanowi odrębną część artykułu;
   - metoda badania, zawierająca: przedmiot i okres badania, źródła danych i zastosowane metody badawcze;
   - wyniki badania;
   - podsumowanie: powinno być zwięzłe i odzwierciedlać istotę problemu badawczego przedstawionego w artykule bez podawania danych liczbowych; wnioski powinny odnosić się do treści artykułu, a w szczególności do celu badań.
7. Bibliografia, zawierająca pełny wykaz prac i materiałów przywołanych w artykule, przygotowana zgodnie z wymogami czasopisma.

Przygotowanie artykułu

1. Tekst należy zapisać alfabetem łacińskim. Nazwy własne, tytuły itp. oryginalnie zapisane innym alfabetem powinny być poddane transliteracji.
2. Nie należy stosować stylów; formatowanie należy ograniczyć do wymogów redakcyjnych.
3. Objętość artykułu łącznie ze streszczeniem, słowami kluczowymi, bibliografią, tablicami, wykresami i innymi materiałami graficznymi nie powinna przekraczać 20 stron maszynopisu.
4. Edytor tekstu: Microsoft Word, format *.doc lub *.docx.
5. Czcionka – Arial, krój prosty:
   - tytuł – 14 pkt, wyśrodkowanie;
   - autor – 12 pkt, wyrównanie do lewej;
   - śródtytuł I stopnia – 14 pkt, wersaliki, wyśrodkowanie;
   - śródtytuł II stopnia – 12 pkt, bold, wyśrodkowanie;
   - tekst główny – 12 pkt;
   - streszczenie, słowa kluczowe i kod JEL – 10 pkt;
   - przypisy – 10 pkt.
6. Marginesy – 2,5 cm z każdej strony.
7. Interlinia – 1,5 wiersza; tablice – 1 wiersz; przed tytułami rozdziałów i podrozdziałów oraz po nich – pusty wiersz.
8. Wcięcie akapitowe – 0,4 cm.

9. Przy wyliczeniach należy posłużyć się listą punktowaną z punktorami w postaci kropek (wysunięcie 0,4 cm, wcięcie 0 cm); wiersze (oprócz ostatniego) zakończone średnikiem.
10. Strony ponumerowane automatycznie.
11. Tablice i elementy graficzne (wykresy, mapy, schematy) muszą być przywołane w tekście.
12. Wykresy, mapy i schematy powinny być zamieszczone w tekście głównym. Dane, na podstawie których opracowano wykresy, powinny być przekazane osobno w pliku programu Excel (lub innym edytowalnym w pakiecie Microsoft Office), ewentualnie wykresy powinny dawać możliwość odczytania z nich danych.
13. Tablice muszą być edytowalne. Nie należy stosować rastrów, cieniowania, pogrubiania czy też podwójnych linii itp.
14. Wskazówki dotyczące opracowywania map znajdują się w publikacji *Mapy statystyczne. Opracowanie i prezentacja danych*, dostępnej na stronie internetowej GUS: https://stat.gov.pl/statystyka-regionalna/publikacje-regionalne/podreczniki-atlasy/podreczniki/mapy-statystyczne-opracowanie-i-prezentacja-danych,1,1.html.
15. Pod tablicami, wykresami, schematami i innymi elementami graficznymi należy podać źródło opracowania.
16. Oznaczenia literowe należy zapisywać następująco: macierze – duże litery, proste, pogrubione (np. $\mathbf{P}$, $\mathbf{N}_{ij}$); wektory – małe litery, kursywa, pogrubione (np. $\boldsymbol{w}$, $\boldsymbol{x}_i$); pozostałe zmienne – małe lub duże litery, kursywa, bez pogrubienia (np. w, x_i, Z).
17. Objaśnienia znaków umownych w tablicach: (–) – zjawisko nie wystąpiło; zero (0) – zjawisko istniało w wielkości mniejszej od 0,5, (0,0) – zjawisko istniało w wielkości mniejszej od 0,05; znak x – wypełnienie pozycji jest niemożliwe lub niecelowe; „w tym” – oznacza, że nie podaje się wszystkich składników sumy.
18. Stosowane są skróty: tablica – tabl., wykres – wykr.
19. Przypisy rzeczowe, słownikowe lub informacyjne należy umieszczać na dole strony. Przypisy bibliograficzne, zgodnie ze standardem APA (American Psychological Association), należy podawać w tekście głównym.
20. Bibliografię należy przygotować zgodnie ze standardem APA.

Zasady przywoływania publikacji w treści artykułu

1. Jeden autor: bez względu na to, ile razy przywoływana jest praca, zawsze należy podać nazwisko autora i datę publikacji pracy, a w przypadku więcej niż jednej pracy danego autora opublikowanej w tym samym roku należy dodać kolejne litery alfabetu przy dacie (np. 2001a). Przykład zapisu: Jak stwierdza Iksiński (2001)... Badania wskazują, że... (Iksiński, 2001).
2. Dwóch autorów: bez względu na to, ile razy przywoływana jest praca, zawsze należy podać nazwiska obu autorów i datę publikacji pracy, a w przypadku więcej niż jednej pracy tych autorów opublikowanej w tym samym roku należy dodać kolejne litery alfabetu przy dacie. Nazwiska autorów zawsze należy łączyć spójnikiem „i”, nawet w przypadku przywoływania publikacji obcojęzycz-

nej. Przykład zapisu: Jak sugerują Iksiński i Nowak (1999)... Badania wskazują, że... (Iksiński i Nowak, 1999).

3. Od trzech do pięciu autorów: przywołanie po raz pierwszy – należy wymienić nazwiska wszystkich autorów, rozdzielając je przecinkami i stawiając spójnik „i" pomiędzy dwoma ostatnimi nazwiskami. Przy kolejnych wskazaniach tej samej pracy należy zastosować określenie „i współpracownicy" (w przypadku umieszczenia przywołania nazwisk w strukturze zdania) lub „i in." (gdy nazwiska autorów nie stanowią części struktury zdania). Przykład zapisu: Przywołanie po raz pierwszy: Jak sugerują Nowak, Iksiński i Jankiewicz (2003)... Badania (Nowak, Iksiński i Jankiewicz, 2003) wskazują, że... Kolejne przywołania: Badania Nowaka i współpracowników (2003)... Badania te wskazują, że (Nowak i in., 2003)...
4. Sześciu i więcej autorów: należy wymienić tylko nazwisko pierwszego autora, zarówno gdy praca przywoływana jest po raz pierwszy, jak i w późniejszych przywołaniach, natomiast pozostałych autorów należy zastąpić określeniem „i współpracownicy" (w przypadku umieszczenia przywołania nazwisk w strukturze zdania) lub „i in." (gdy nazwiska nie stanowią części struktury zdania). W literaturze załącznikowej należy umieścić nazwiska wszystkich autorów pracy. Przykład zapisu: Nowakowski i współpracownicy (1997) twierdzą, że... Pierwsze badania na ten temat (Nowakowski i in., 1997) sugerują...
5. Przywoływanie jednocześnie kilku prac: należy wymienić je alfabetycznie według nazwiska pierwszego autora. Przywołania kolejnych prac muszą być oddzielone średnikiem. Lata wydania prac tego samego autora/autorów muszą być oddzielone przecinkiem. Przykład zapisu: Iksiński (2001); Nowak i Iksiński (1999, 2005); (Iksiński, 1997, 1999, 2004a, 2004b; Nowak i Iksiński, 1999).
6. Przywoływanie pracy za innym autorem: stosuje się w tekście, natomiast w bibliografii należy umieścić jedynie pracę czytaną. Przykład zapisu: Jak wykazał Nowakowski (1990; za: Zieniecka, 2007)... Badania sugerują, że... (Nowakowski, 1990; za: Zieniecka, 2007).
7. Bibliografia powinna być zamieszczona na końcu opracowania. Prace należy zapisać alfabetycznie według nazwiska pierwszego autora. W przypadku dwóch lub więcej prac tego samego autora/autorów należy je uporządkować według roku publikacji. Jeśli kilka prac tego samego autora/autorów zostało opublikowanych w tym samym roku, należy uporządkować prace alfabetycznie według tytułu i wstawić litery a, b, c itd. po roku publikacji. Zapis dotyczący każdej nowej pracy należy zacząć bez wcięcia, wyrównanie do lewego marginesu, a w kolejnych wierszach danego zapisu stosować wcięcie 0,4 cm.

Przykłady opisu bibliograficznego

1. Artykuł w czasopiśmie, w którym każdy kolejny numer/zeszyt (issue) w ramach jednego rocznika ma osobną numerację stron (w każdym zeszycie pierwsza strona opatrzona jest numerem 1): Nazwisko, X., Nazwisko 2, X. Y.,

Nazwisko 3, Z. (rok). Tytuł artykułu. *Tytuł czasopisma*, *rocznik*(zeszyt), strona początku–strona końca.

2. Artykuł w czasopiśmie, w którym kolejne numery/zeszyty (issues) w ramach jednego rocznika nie mają osobnej numeracji stron (pierwsza strona w kolejnym zeszycie opatrzona jest numerem kolejnym po ostatniej stronie w zeszycie poprzednim): Nazwisko, X., Nazwisko 2, X. Y., Nazwisko 3, Z. (rok). Tytuł artykułu. *Tytuł czasopisma*, *rocznik*, strona początku–strona końca. Jeśli artykuł ma numer DOI (Digital Object Identifier), należy podać go na końcu opisu bibliograficznego: Nazwisko, X., Nazwisko 2, X. Y. (rok). Tytuł artykułu. *Tytuł czasopisma*, *rocznik*, strona początku–strona końca. DOI: xxxxx.
3. Książka: Nazwisko, X., Nazwisko 2, X. Y. (rok). *Tytuł książki*. Miejsce wydania: wydawnictwo.
4. Książka napisana pod redakcją: Nazwisko, X. (red.). (rok). *Tytuł książki*. Miejsce wydania: wydawnictwo.
5. Rozdział w pracy zbiorowej: Nazwisko, X. (rok). Tytuł rozdziału. W: Y. Nazwisko, B. Nazwisko 2 (red.), *Tytuł książki* (s. strona początku–strona końca). Miejsce wydania: wydawnictwo.
6. Jeśli dany tekst znajduje się na stronie internetowej i nie jest artykułem w czasopiśmie, książką ani rozdziałem w książce, należy podać autora, datę publikacji (jeśli jest znana), tytuł, a następnie zamieścić informacje o stronie, z której został pobrany, oraz – jeśli są to materiały informacyjne – datę dostępu. Tekst: Nazwisko, X. (rok). *Tytuł tekstu*. Pobrane z: adres strony internetowej (dostęp: 21.03.2019).

Artykuł przygotowany w sposób niezgodny z powyższymi wskazówkami będzie odesłany z prośbą o dostosowanie jego formy do wymagań redakcji.

Zakres tematyczny działów
Thematic scope of sections

(for information go to: ws.stat.gov.pl/AimScope)

STUDIA METODOLOGICZNE

W tym dziale zamieszczane są artykuły naukowe przedstawiające teoretyczne rozwiązania metodologiczne ze wskazaniem ich praktycznej użyteczności, w tym prace przeglądowe i porównawcze oraz dotyczące etyki w statystyce. Poruszane w nich zagadnienia obejmują różne dziedziny statystyki, ekonomii matematycznej i ekonometrii. Omawiane rezultaty badawcze mogą znaleźć efektywne zastosowanie w badaniach empirycznych oraz analizach statystycznych i służyć podnoszeniu ich jakości, jak również powiększeniu zasobu informacyjnego.

STATYSTYKA W PRAKTYCE

Dział ten zawiera artykuły poświęcone nowatorskim zastosowaniom w praktyce znanych narzędzi i modeli statystycznych oraz analizie i ocenie statystycznej zjawisk społeczno-ekonomicznych i innych; zamieszczane tu prace opierają się w szczególności na danych pochodzących z zasobów statystyki publicznej. Zastosowania w praktyce obejmują również wykorzystanie narzędzi informatycznych do uzyskiwania i przetwarzania informacji statystycznych, naliczania danych wynikowych, ich prezentacji i rozpowszechniania. Może to też dotyczyć opracowań stosujących nowoczesne techniki programistyczne pozwalające na efektywną komunikację z systemami informacyjnymi oraz ułatwiające wykorzystanie danych wynikowych. Publikowane są także artykuły sygnalizujące problemy związane z projektowaniem badań statystycznych, uzyskiwaniem, integracją i przetwarzaniem danych oraz generowaniem wynikowych informacji statystycznych i kontrolą ich ujawniania wraz z propozycjami efektywnych rozwiązań w tym zakresie.

STUDIA INTERDYSCYPLINARNE. WYZWANIA BADAWCZE

To blok tematyczny zawierający artykuły wskazujące i podejmujące wyzwania badawcze, które są szczególnie istotne ze względu na rosnące potrzeby współczesnych użytkowników danych statystycznych i wymagają zaangażowania znacznych nakładów pracy, środków oraz rozwiązań z różnych dziedzin nauki i techniki. W dziale tym publikowane są również opracowania dotyczące: wykorzystania technologii informacyjnych i komunikacyjnych (ICT), gospodarki opartej na wiedzy, problematyki innowacyjności, przepływu informacji we współczesnym społeczeństwie oraz przetwarzania i analizy zagadnień związanych z data science i big data, a zatem problematyki bardzo często powiązanej z działaniami interdyscyplinarnymi.

EDUKACJA STATYSTYCZNA

W tym dziale zamieszczane są artykuły dotyczące metod i efektów nauczania statystyki oraz popularyzacji myślenia statystycznego. Odnosi się to zwłaszcza do problemów związanych z kształceniem w zakresie umiejętności stosowania statystyki na wszystkich poziomach edukacji, a także do wykorzystywania nowoczesnych koncepcji i metod dydaktycznych oraz pomocy naukowych w nauczaniu statystyki. Uwaga skoncentrowana jest na rozumieniu prawdopodobieństwa i statystyki, badaniach z zakresu nauczania statystyki, postaw i zachowań społecznych w odniesieniu do tej dziedziny wiedzy, jak również na rozumieniu informacji statystycznych. Ponadto ukazywane są problemy związane z prezentacją danych statystycznych oraz ich interpretacją w powszechnym obiegu informacyjnym, np. w środkach społecznego przekazu.

Z DZIEJÓW STATYSTYKI

Prace publikowane w tym dziale poświęcone są historii prowadzenia obserwacji statystycznych oraz rozwoju ich metodologii i narzędzi. Ponadto zamieszczane są tu informacje dotyczące życia i osiągnięć zawodowych wybitnych statystyków, jak również najważniejszych instytucji i organizacji statystycznych w Polsce i za granicą.

INFORMACJE. RECENZJE. DYSKUSJE

Jedyny dział zawierający teksty nierecenzowane i niemające charakteru artykułów naukowych. Obejmuje informacje o najważniejszych wydarzeniach dotyczących statystyki polskiej i międzynarodowej, a także sprawozdania z konferencji naukowych, recenzje książek i opracowań z zakresu statystyki i jej zastosowań, rekomendacje nowych, istotnych i ciekawych pozycji wydawniczych z tego obszaru wiedzy, jak również odpowiedzi autorów na recenzje oraz polemiki, dyskusje i sprostowania dotyczące artykułów zamieszczonych na łamach czasopisma.