

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  
PAŹDZIERNIK 2019  
OCTOBER

Numer **10** (701)  
Issue



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  
PAŹDZIERNIK 2019  
OCTOBER

Numer **10** (701)  
Issue

---

---

## RADA NAUKOWA / SCIENTIFIC COUNCIL

dr Dominik Rozkrut (przewodniczący/chairman) – Uniwersytet Szczeciński, prof. Anthony Arundel – University of Tasmania in Hobart, dr hab. Bożena Balcerzak-Paradowska, prof. IPiSS – Instytut Pracy i Spraw Socjalnych, prof. Eric Bartelsman – Vrije Universiteit Amsterdam, prof. dr hab. Czesław Domański – Uniwersytet Łódzki, dr hab. Elżbieta Gołata, prof. UEP – Uniwersytet Ekonomiczny w Poznaniu, prof. Semen Matkovskiy, PhD – Ivan Franko National University of Lviv, prof. dr hab. Włodzimierz Okrasa – Uniwersytet Kardynała Stefana Wyszyńskiego w Warszawie, prof. dr hab. Józef Oleński – Uczelnia Łazarskiego, prof. dr hab. Tomasz Panek – Szkoła Główna Handlowa w Warszawie, prof. Juan Manuel Rodríguez Poo – University of Cantabria, assoc. prof. ing. Iveta Stankovičová, PhD – Comenius University in Bratislava, prof. dr hab. Marek Walesiak – Uniwersytet Ekonomiczny we Wrocławiu, prof. dr hab. Józef Zegar – Instytut Ekonomiki Rolnictwa i Gospodarki Żywnościowej – Państwowy Instytut Badawczy

sekretarz/secretary: Paulina Kucharska-Singh

---

## KOLEGIUM REDAKCYJNE / EDITORIAL BOARD

prof. Tudorel Andrei – Bucharest Academy of Economic Studies, mgr Renata Bielak – Główny Urząd Statystyczny, dr Marek Cierpień-Wolan – Uniwersytet Rzeszowski, dr hab. Grażyna Dehnel, prof. UEP – Uniwersytet Ekonomiczny w Poznaniu, dr Jacek Kowalewski – Uniwersytet Ekonomiczny w Poznaniu, dr Jan Kubacki – Urząd Statystyczny w Łodzi, mgr Władysław Wiesław Łagodziński – Polskie Towarzystwo Statystyczne, dr Grażyna Marciniak – Główny Urząd Statystyczny, dr hab. Andrzej Młodak – Państwowa Wyższa Szkoła Zawodowa im. Prezydenta Stanisława Wojciechowskiego w Kaliszu, dr Stanisław Paradysz, dr hab. Mateusz Piepień – Uniwersytet Ekonomiczny w Krakowie, ing. Marek Rojiček, PhD – University of Economics Prague, assoc. prof. Anna Shostya, PhD – Pace University in New York, dr hab. Małgorzata Tarczyńska-Łuniewska, prof. US – Uniwersytet Szczeciński, dr Wioletta Wrzaszcz – Instytut Ekonomiki Rolnictwa i Gospodarki Żywnościowej – Państwowy Instytut Badawczy, dr inż. Agnieszka Zgierska – Główny Urząd Statystyczny

## ZESPÓŁ REDAKCYJNY / EDITORIAL STAFF

redaktor naczelny / editor-in-chief: Marek Cierpień-Wolan

zastępca redaktora naczelnego / deputy editor-in-chief: Andrzej Młodak

redaktorzy tematyczni / thematic editors: Jan Kubacki, Małgorzata Tarczyńska-Łuniewska, Agnieszka Zgierska

redaktor merytoryczny / substantive editor: Wioletta Wrzaszcz

redaktorzy językowi / language editors: Ewa Antoniak, Xawery Stańczyk, Małgorzata Zygmunt (język polski), Waleria Wrennall (język angielski)

sekretarz/secretary: Małgorzata Zygmunt

---

## 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](http://ws.stat.gov.pl)**

**An electronic edition of the journal is an original one. It is available at [ws.stat.gov.pl](http://ws.stat.gov.pl)**

© Copyright by Główny Urząd Statystyczny / Statistics Poland



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](http://www.prenumerata.ruch.com.pl)

---

## Spis treści Contents

<b>OD REDAKCJI</b> .....	<b>4</b>
<b>FROM THE EDITORIAL TEAM</b>	
<b>STUDIA METODOLOGICZNE</b>	
<b>METHODOLOGICAL STUDIES</b>	
Maciej Beręsewicz, Grzegorz Gudaszewski, Marcin Szymkowiak	
Estymacja liczby cudzoziemców w Polsce z wykorzystaniem metody capture- -recapture .....	<b>7</b>
Estimation of the number of foreigners in Poland using the capture-recapture method	
<b>STATYSTYKA W PRAKTYCE</b>	
<b>STATISTICS IN PRACTICE</b>	
Sergiusz Herman	
Impact of joint-stock companies' financial condition on real activities manipulation to manage earnings .....	<b>36</b>
Wpływ kondycji finansowej spółek akcyjnych na podejmowanie działań o charak- terze realnym w celu zarządzania zyskami	
Andrzej Ochocki	
Starzenie się ludności a redystrybucja dochodów .....	<b>53</b>
Population ageing and income redistribution	
<b>Z DZIEJÓW STATYSTYKI</b>	
<b>FROM THE HISTORY OF STATISTICS</b>	
Bożena Łazowska	
Spis ludności na ziemiach wschodnich Rzeczypospolitej Polskiej w 1919 r. ....	<b>63</b>
Census of the Eastern Lands of the Second Polish Republic in 1919	
<b>INFORMACJE. RECENZJE. DYSKUSJE</b>	
<b>INFORMATION. REVIEWS. DISCUSSIONS</b>	
Zenon Wiśniewski	
Miroslaw Szreder <i>Zrozumieć świat liczb</i> (Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk 2019) .....	<b>77</b>
Miroslaw Szreder <i>To understand the world of numbers</i> (Gdańsk University Press, Gdańsk 2019)	
Justyna Gustyn	
Wydawnictwa GUS. Wrzesień 2019 .....	<b>81</b>
Publications of Statistics Poland. September 2019	
<b>DLA AUTORÓW</b> .....	<b>83</b>
<b>FOR THE AUTHORS</b>	
<b>ZAKRES TEMATYCZNY DZIAŁÓW</b> .....	<b>93</b>
<b>THEMATIC SCOPE OF SECTIONS</b>	

## Od redakcji

W październikowym numerze „Wiadomości Statystycznych. The Polish Statistician” zamieszczamy artykuły dotyczące rozwiązań metodologicznych, zastosowań wybranych narzędzi statystycznych w praktyce oraz opracowanie poświęcone historii statystyki.

Wydanie otwiera artykuł Macieja Beręsewicza, Grzegorza Gudaszewskiego i Marcina Szymkowiaka *Estymacja liczby cudzoziemców w Polsce z wykorzystaniem metody capture-recapture*, prezentowany w dziale Studia metodologiczne. Badacze przedstawiają metodę capture-recapture oraz wyniki przeprowadzonego szacunku populacji cudzoziemców przebywających w Polsce pod koniec lat 2015 i 2016. W badaniu posłużono się modelami logarytmiczno-liniowymi z szeregiem zmiennych pomocniczych. Praca ta wpisuje się w nurt badań nad populacjami trudnymi do zbadania, czyli charakteryzującymi się brakiem wyczerpującego operatu losowania oraz problemami w uzyskaniu informacji od jednostek do nich należących.

W dziale Statystyka w praktyce publikujemy dwa artykuły. Pierwszy to *Impact of joint-stock companies' financial condition on real activities manipulation to manage earnings* Sergiusza Hermana. Autor bada aktywność spółek będących w gorszej kondycji finansowej w zakresie zarządzania zyskami. Analizuje działania o charakterze realnym podejmowane w celu zarządzania zyskami w zależności od rodzaju działalności gospodarczej przedsiębiorstwa i rozważa zagadnienie manipulacji danymi o sprzedaży, produkcji oraz kosztach pośrednich przedsiębiorstw.

Z kolei artykuł Andrzeja Ochockiego *Starzenie się ludności a redystrybucja dochodów* przedstawia uwarunkowania dynamiki wynagrodzeń i emerytur w Polsce w latach 2010–2017 wraz z prognozą studialną do 2025 r. Podjęte zagadnienie badawcze jest szczególnie ważne w kontekście obserwowanych zmian demograficznych. W opisanym w pracy badaniu zastosowano autorski model międzypokoleniowej redystrybucji dochodów.

W dziale Z dziejów statystyki Bożena Łazowska w artykule *Spis ludności na ziemiach wschodnich Rzeczypospolitej Polskiej w 1919 r.* omawia organizację oraz wyniki spisu przeprowadzonego w grudniu 1919 r. na terenach administrowanych przez Zarząd Cywilny Ziemi Wschodnich. Autorka przybliży zadania Zarządu, okoliczności przeprowadzenia spisu oraz jego znaczenie dla późniejszych badań Głównego Urzędu Statystycznego.

Numer zamykają omówienie wybranych nowości wydawniczych Głównego Urzędu Statystycznego oraz recenzja książki Mirosława Szredera *Zrozumieć świat liczb*, wydanej w tym roku nakładem Uniwersytetu Gdańskiego. Omawiana publikacja składa się z 27 tekstów naukowych i popularnonaukowych, w większości drukowanych wcześniej w prasie, które prezentują możliwości zastosowań metod statystycznych. W ocenie recenzenta Zenona Wiśniewskiego zaletą książki jest jej przystępność, zachęcająca do poznania warsztatu badawczego statystyków.

Zapraszamy do lektury.

## From the editorial team

The October issue of *Wiadomości Statystyczne. The Polish Statistician* has articles focused on methodological solutions, applications of selected statistical tools in practice, and the history of statistics.

The issue opens with Maciej Beręsewicz, Grzegorz Gudaszewski and Marcin Szymkowiak's article entitled *Estimation of the number of foreigners in Poland using the capture-recapture method*, from the Methodological studies section. The authors present the capture-recapture method and the results of the estimation of the number of foreigners staying in Poland at the end of 2015 and 2016. They use log-linear models with a series of auxiliary variables. The study fits in the trend of investigating hard-to-survey populations, where there is no exhaustive sampling frame and it is problematic to obtain information from the members.

The following two articles, which come under the Statistics in practice section, are *Impact of joint-stock companies' financial condition on real activities manipulation to manage earnings* by Sergiusz Herman, written in English, and *Population ageing and income redistribution* by Andrzej Ochocki. The former article examines the activity of companies in poor financial condition regarding earnings management. More specifically, the author analyses types of real activities undertaken in particular business sectors in order to manage earnings. The paper also focuses on the manipulation of different types of data, such as sales and production volumes and companies' indirect costs.

The latter article, by Andrzej Ochocki, presents the determinants of the dynamics of remuneration and pension growth in Poland in 2010–2017 and the study forecast until 2025. The research problem discussed here is especially important in the context of the observed demographic changes. The analysis has been performed according to the author's original Intergenerational Income Transfer Model.

The article entitled *Census of the Eastern Lands of the Second Polish Republic in 1919* by Bożena Łazowska, presented in the From the history of statistics section, describes the process of the organization and the results of a census carried out in December 1919 in the territories controlled by the Civil Administration of the Eastern Lands. The author presents the Administration's tasks, the circumstances in which the census was performed, and explains its significance for all the research that has been subsequently carried out by Statistics Poland.

The issue concludes with brief descriptions of a selection of Statistics Poland's latest publications and a review of Mirosław Szreder's book *To understand the world of numbers*, published by the Gdańsk University. The reviewed volume consists of 27 scientific and popular science papers, most of which have already been published in the press. They all show different applications of statistical methods. According to reviewer Zenon Wiśniewski, one of the book's biggest merits is its approachability, which encourages readers to familiarise themselves with the tools and methods used by statisticians.

We wish you a pleasant read.



## Estymacja liczby cudzoziemców w Polsce z wykorzystaniem metody capture-recapture<sup>1</sup>

Maciej Beręsewicz<sup>a</sup> , Grzegorz Gudaszewski<sup>b</sup> , Marcin Szymkowiak<sup>a</sup> 

**Streszczenie.** Celem artykułu jest przedstawienie metody badań oraz wyników szacunku populacji cudzoziemców przebywających w Polsce. W badaniu wykorzystano administracyjne źródła danych. Zastosowano metodę capture-recapture bazującą na modelach log-liniowych. Szacuje się, że w 2015 i 2016 r. na terenie Polski mogło przebywać odpowiednio ok. 500 tys. (przyjmując 95-procentowy przedział ufności – od 369 tys. do 724 tys.) oraz ok. 744 tys. (od 601 tys. do 943 tys.) cudzoziemców. Jest to pierwsza tego typu kompleksowa analiza dotycząca próby estymacji liczby cudzoziemców w Polsce, która wpisuje się w nurt badań nad populacjami trudnymi do zbadania. Należy jednak mieć na uwadze konieczność spełnienia założeń tej metody, co również stanowił przedmiot rozważań autorów.

**Słowa kluczowe:** estymacja liczby cudzoziemców, populacja trudna do zbadania, metoda capture-recapture, analiza log-liniowa, rejestry administracyjne

## Estimation of the number of foreigners in Poland using the capture-recapture method

**Abstract.** The aim of this paper is to present the methodology and the results of the estimation of the number of foreigners staying in Poland. Administrative data sources were used in the research. The authors adopted the capture-recapture method based on log-linear models. As a result, the number of foreigners staying in Poland in 2015 and 2016 has been estimated at around 500,000 persons (95% CI: 369,000–724,000) and around 744,000 persons (601,000–943,000), respectively. The study is the first comprehensive analysis of this kind which aims at estimating the number of foreigners in Poland, and thus fits in the current of research on hard-to-survey populations. It has to be remembered, however, that capture-recapture method requires close observance of its strict rules in order to be effective, which is also discussed in depth in the paper.

**Keywords:** estimation the number of foreigners in Poland, hard-to-survey population, capture-recapture method, log-linear analysis, administrative registers

**JEL:** C81, J61, J68, F22

---

<sup>1</sup> Artykuł został przygotowany na podstawie raportu podsumowującego pracę badawczą *Cudzoziemcy na krajowym rynku pracy w ujęciu regionalnym* zrealizowaną w ramach projektu *Wsparcie systemu monitorowania polityki spójności w perspektywie finansowej 2014–2020 oraz programowania i monitorowania polityki spójności po 2020 roku* współfinansowanego przez Unię Europejską ze środków Programu Operacyjnego Pomoc Techniczna 2014–2020.

<sup>a</sup> Uniwersytet Ekonomiczny w Poznaniu, Instytut Informatyki i Ekonomii Ilościowej.

<sup>b</sup> Główny Urząd Statystyczny, Departament Badań Demograficznych.



Coraz częściej, zarówno na szczeblu rządowym, samorządowym, jak i lokalnym, porusza się kwestię konieczności ustalenia rzeczywistej liczby cudzoziemców przebywających w Polsce stale i czasowo oraz podejmujących tu pracę. Szczególnie istotną informację dla realizowania polityki ludnościowej, migracyjnej i gospodarczej stanowią cechy demograficzno-społeczne i ekonomiczne cudzoziemców. Ważne jest także określenie skali imigracji nierejestrowanej, tj. pozostającej poza ewidencją. Obecnie nie ma w Polsce miarodajnego i bezpośredniego źródła, które dostarczałoby wiarygodnych danych na ten temat. Należy również zwrócić uwagę, że imigracja cudzoziemców jest zjawiskiem zróżnicowanym przestrzennie, zwłaszcza w kontekście regionalnych rynków pracy.

Pozyskanie danych dotyczących liczby cudzoziemców, w tym nierejestrowanych imigrantów, to dla służb statystyki publicznej w Polsce istotne wyzwanie metodologiczne. Po pierwsze, rejestry administracyjne dostarczają informacji o populacji *de iure* (zarejestrowanej), podczas gdy statystyka zainteresowana jest populacją *de facto* (zarejestrowani i niezarejestrowani). Po drugie, cudzoziemcy są populacją trudną do zbadania przy wykorzystaniu tradycyjnych metod statystycznych. Populacja taka charakteryzuje się bowiem brakiem dostępnego (wyczerpującego) operatu losowania oraz trudnością w pozyskaniu informacji od jednostek do niej należących. O ile rozpoznanie problemów występujących w populacjach trudnych do zbadania jest możliwe na gruncie badań statystycznych (można zastosować przykładowo dobór jednostek do badania, opierając się na metodzie kuli śnieżnej i jej rozszerzeniu – metodzie RDS<sup>2</sup>), o tyle proces estymacji wielkości takiej zbiorowości jest z metodologicznego punktu widzenia poważnym wyzwaniem badawczym. W literaturze przedmiotu opisano jednak odpowiednie metody statystyczne, które umożliwiają estymację wielkości populacji trudnych do zbadania, bazujące na technikach capture-recapture (Böhning, van der Heijden i Bunge, 2017). Należą do nich rozwiązania, w których wykorzystuje się jedno (Godwin i Böhning, 2017; van der Heijden, Cruyff i van Houwelingen, 2003) albo co najmniej dwa źródła danych (van der Heijden, Whittaker, Cruyff, Bakker i van der Vliet, 2012; Zhang, 2008). Skuteczność tych technik w praktyce w dużej mierze zależy od dostępności danych statystycznych i jest uwarunkowana koniecznością spełnienia odpowiednich założeń leżących u podstaw poszczególnych metod.

Celem artykułu jest przedstawienie metody badań oraz wyników szacunku populacji cudzoziemców przebywających w Polsce. Podjęto w nim próbę osza-

---

<sup>2</sup> Respondent Driven Sampling – metoda doboru jednostek do próby sterowana przez respondentów. Jest to zmodyfikowana wersja metody kuli śnieżnej, w której stosuje się podwójny system zachęt polegający na wynagrodzeniu respondenta za wzięcie udziału w badaniu i zwerbowaniu kolejnych osób, które biorą w nim udział. W metodzie RDS wykorzystuje się informacje na temat sieci powiązań osób należących do danej zbiorowości.

cowania wielkości populacji cudzoziemców przebywających w Polsce pod koniec 2015 i 2016 r. i według kraju obywatelstwa<sup>3</sup>. W tym celu posłużono się odpowiednio zbudowanym modelem log-liniowym z szeregiem zmiennych pomocniczych. Przyjęto, zgodnie z Ustawą z dnia 12 grudnia 2013 r. o cudzoziemcach (Dz.U. 2016 poz. 1990 z późn. zm.), że cudzoziemcem jest osoba nieposiadająca obywatelstwa polskiego lub bezpaństwowiec. Należy zaznaczyć, że wybór okresu oraz źródeł danych był podyktowany ich dostępnością dla statystyki publicznej w ramach programu badań statystycznych statystyki publicznej (PBSSP). Pozyskanie innych danych jednostkowych było niemożliwe. Udostępnione dane przedstawiają stan na 31 grudnia 2016 r. Nie oznacza to, że w tym dniu wszyscy badani cudzoziemcy przebywali na terenie Polski, podobnie jak w przypadku rejestru PESEL, który nie gwarantuje, że na terenie Polski w danym dniu przebywa określona liczba obywateli polskich.

## PRZEGLĄD LITERATURY

### Pojęcie populacji trudnej do zbadania

W literaturze przedmiotu populacje trudne do zbadania są definiowane na różne sposoby (Tourangeau, Edwards, Johnson, Wolter i Bates, 2014). Ze względu na to, że w wielu badaniach częściowych mamy do czynienia z dużym odsetkiem odmów, terminem tym można byłoby określić każdą z badanych populacji. Jednak pojęcie populacji trudnej do zbadania ma inne znaczenie i odnosi się do zbiorowości, których badanie wiąże się ze szczególnymi wyzwaniami metodologicznymi różnego rodzaju i jest bardziej skomplikowane w porównaniu z badaniem innych populacji. Są to np. populacje rzadkie, ukryte, z których jednostkami trudno nawiązać kontakt i współpracować.

Mówiąc o populacjach trudnych do zbadania, należy rozróżnić populacje (Tourangeau i in., 2014):

- z których trudno wylosować jednostki do próby (ang. *hard-to-sample*) – bardzo rzadko zdarza się, aby istniał właściwy operat losowania, z którego można byłoby wylosować jednostki do próby, wykorzystując odpowiedni schemat jej pobierania. Z tego względu stosuje się nielosowe doборы jednostek do próby, wśród których szczególną rolę odgrywa wspomniana już metoda kuli śnieżnej czy doboru sterowana przez respondenta. Można również zastoso-

---

<sup>3</sup> W pracy badawczej, na podstawie której powstał niniejszy artykuł, rozważany był również poziom województw i podregionów. Opracowano także podstawową charakterystykę cudzoziemców uwzględniającą wybrane cechy demograficzno-społeczne, obywatelstwo czy status na rynku pracy na podstawie danych z Narodowego Spisu Powszechnego Ludności i Mieszkań 2011, Badania Aktywności Ekonomicznej Ludności oraz zezwoleń i oświadczeń Ministerstwa Rodziny, Pracy i Polityki Społecznej, które nie będą jednak omawiane w tym miejscu.

wać inne techniki doboru jednostek, zwłaszcza w odniesieniu do populacji rzadkich i trudno uchwytnych, takie jak losowanie odwrotne, lokacyjne czy schematy linia-przecięcie oraz śledzenia łączy (Jędrzejczak i Kubacki, 2014). Jednak nawet w takim przypadku dobór jednostek do próby jest problematyczny, gdyż omawiane populacje mogą być mobilne bądź nieuchwytne. Przykład: osoby bezdomne lub pracujący cudzoziemcy;

- których jednostki trudno zidentyfikować (ang. *hard-to-identify*) – w niektórych przypadkach, zwłaszcza stygmatyzowanych grup społecznych, członkowie populacji mogą nie chcieć udostępnić swoich danych z obawy przed ujawnieniem nielegalnego lub wstydlwego statusu społecznego, co uniemożliwia identyfikację jednostek należących do takich populacji. Przykład: narkomani, alkoholicy, mniejszości (LGBT, wyznawcy określonych religii czy ideologii);
- których jednostki trudno odnaleźć i z którymi trudno nawiązać kontakt (ang. *hard-to-find-and-contact*) – nawiązanie kontaktu komplikuje przede wszystkim mobilność tego typu populacji. Przykład: niezameldowani cudzoziemcy, członkowie kultur koczowniczych (Beduini z południowo-zachodniej Azji czy Tuaregowie z Afryki Północnej), mniejszości wędrownie (Romowie w Europie), osoby bezdomne;
- których jednostki trudno namówić do wzięcia udziału w badaniu (ang. *hard-to-persuade*) – niechęć do udziału w badaniu może wynikać z drażliwości poruszanej tematyki bądź z braku czasu. Przykład: aktywni zawodowo, pracujący w szarej czy czarnej strefie, cudzoziemcy;
- których jednostki można zachęcić do wzięcia udziału w badaniu, ale trudno z nimi przeprowadzić wywiad (ang. *hard-to-interview*) – przeszkodą w przeprowadzeniu wywiadu może być konieczność uzyskania zgody przełożonego, opiekuna prawnego czy rodzica na udział danej jednostki w badaniu, np. w przypadku osoby z niepełnosprawnością lub nieposługującej się językiem, w którym przygotowano kwestionariusz, a także gdy badanie należy przeprowadzić w obszarze konfliktu zbrojnego. Przykład: więźniowie, osoby z niepełnosprawnością intelektualną albo psychiczną czy cudzoziemcy nieznający języka danego kraju.

Jak pokazują powyższe rozważania, problemy związane z badaniem określonych populacji mogą wynikać z wielu czynników. Tak jest w przypadku cudzoziemców w Polsce, spośród których trudno wylosować jednostki do próby (brak pełnego operatu losowania oraz kompleksowych źródeł danych statystycznych, z których można czerpać wiedzę na temat cudzoziemców), z którymi trudno nawiązać kontakt (mobilność cudzoziemców na rynku pracy oraz brak stałego miejsca zamieszkania) i przeprowadzić wywiad ze względu na barierę językową. Czynniki te powodują również, że estymacja liczebności tego typu populacji, zwłaszcza z uwzględnieniem dodatkowych przekrojów, jest niezwykle złożonym zadaniem. W literaturze przedmiotu proponuje się jednak pewne rozwiązania,

które mogą być remedium na problemy związane z określeniem rzeczywistych rozmiarów populacji trudnych do zbadania. Należą one do grupy technik określanych terminem capture-recapture<sup>4</sup>.

### **Szacowanie wielkości populacji trudnej do zbadania z wykorzystaniem metody capture-recapture**

Metoda capture-recapture wywodzi się z nauk przyrodniczych. Pierwotnie użyto jej do oszacowania liczby ryb w jeziorze (Goudie i Goudie, 2007). Idea tego podejścia polega na tym, że w typowym badaniu z obszaru nauk przyrodniczych przeprowadzonym metodą capture-recapture na analizowanym terytorium umieszcza się pułapki lub siatki w celu wielokrotnego wylapywania osobników danej populacji. W pierwszej próbie zostaje złowiona pewna liczba osobników, które po oznakowaniu są wypuszczane na wolność. W każdej kolejnej próbie zapisuje się i znakuje każde nieoznaczone zwierzę, notuje się każde zwierzę, które zostało wcześniej oznakowane, i ponownie wypuszcza się je na wolność. Po zakończeniu badania uzyskuje się pełną historię złowień dla każdego osobnika. Badania tego typu określane są mianem mark-recapture, tag-recapture czy multiple-record system.

W najprostszej wersji metoda capture-recapture składa się z dwóch prób lub źródeł<sup>5</sup>. Pierwsza to próba zawierająca osobniki złowione za pierwszym razem, druga – zwierzęta złowione za drugim razem. Ten szczególny przypadek złożony z dwóch prób w kontekście szacowania błędu niedostatecznego pokrycia określane jest jako system podwójny (ang. *dual system*) lub system podwójnego zapisu (ang. *dual-system record*). Od wielu lat metodę wielokrotnych złowień stosuje się do szacowania parametrów demograficznych w populacjach zwierzęcych. Biolodzy już dawno zauważyli, że nie jest konieczne ani nawet możliwe zliczenie wszystkich zwierząt w celu dokładnego oszacowania wielkości populacji. Informacja na temat liczby ponownych złowień (lub proporcji ponownych złowień) uzyskiwana poprzez znakowanie odgrywa tu istotną rolę, ponieważ można ją wykorzystać do oszacowania liczby osobników nieujętych w próbach, przyjmując odpowiednie założenia. W najprostszym ujęciu można założyć, że w przypadku gdy liczba ponownie złowionych osobników w kolejnych próbach jest niewielka, rozmiar populacji jest większy niż liczba unikatowych osobników, które zostały złowione. Natomiast jeśli wskaźnik ponownych złowień jest stosunkowo wysoki, można przypuszczać, że złowiono większość zwierząt

---

<sup>4</sup> Autorzy postępują się angielskim terminem *capture-recapture*, który w dosłownym tłumaczeniu na język polski mógłby brzmieć „metoda wielokrotnego połowu”. Określenie to nie oddaje jednak istoty tego podejścia, w szczególności w odniesieniu do rejestrów administracyjnych, w których nie dokonano losowań czy „połowów” jednostek.

<sup>5</sup> Możliwe jest również zastosowanie metody w przypadku jednego źródła, o czym będzie mowa później.

z danej populacji. Pomysł zastosowania techniki złożonej z dwóch prób można odnaleźć w pracach Pierre'a Simona de Laplace'a z 1786 r., który wykorzystał ją do szacowania liczby ludności Francji w 1802 r., a nawet wcześniej, w pracach Johna Graunta, który za jej pomocą oszacował skutki zarazy wśród ludności Anglii ok. 1600 r. W dziedzinie ekologii techniki tej najwcześniej użyto w badaniach Petersena i Dahla dotyczących populacji ryb odpowiednio w 1896 i 1907 r. oraz w przeprowadzonym przez Lincolna w 1930 r. badaniu powrotów zaobrazkowanych ptaków wodnych. Modele oparte na dwóch próbach zostały rozszerzone na przypadki zawierające większą liczbę prób przez Schnabela w 1938 r., stąd też metoda wielokrotnego połowu nazywana jest również spisem Schnabela. Bardziej zaawansowana teoria statystyczna i procedury wnioskowania w kontekście populacji trudnych do zbadania pojawiły się po publikacji prac Darrocha, który opracował zagadnienie pod względem matematycznym (Böhning, van der Heijden i Bunge, 2018).

Założenia stosowane w odniesieniu do populacji zwierzęcych klasyfikuje się generalnie jako modele zamknięte i otwarte. W modelu zamkniętym zakłada się, że wielkość populacji będącej przedmiotem badania jest stała w okresie prowadzenia badania. Założenie to jest zwykle zachowane w przypadku danych zbieranych w stosunkowo krótkim czasie poza okresem godowym. W modelu otwartym dopuszcza się przyrosty (narodziny lub imigracja) lub ubytki (śmierć lub emigracja) w populacji. Założenie otwartej populacji zwykle wykorzystuje się w długoterminowych badaniach zwierząt lub ptaków wędrownych. Poza wielkością populacji w momencie poszczególnych prób badane parametry obejmują również wskaźnik przeżywalności oraz liczbę narodzin pomiędzy próbami.

Warto zaznaczyć, że współcześnie pojęcie capture-recapture jest szerokie i odnosi się do wielu metod mających na celu oszacowanie wielkości nieznannej populacji. Zwykle wykorzystuje się różnego rodzaju narzędzia statystyczne, np. modele log-liniowe, modele klas ukrytych czy uogólnione modele liniowe. W prezentowanym badaniu zastosowano metodę capture-recapture wykorzystującą analizę log-liniową. Natomiast warto pamiętać, że wybór odpowiedniej techniki w estymacji liczebności populacji trudnych do zbadania podyktowany jest w dużej mierze liczbą dostępnych źródeł, którą można podzielić na przypadki wyłącznie jednego albo dwóch lub więcej źródeł.

Kluczowym aspektem metody capture-recapture są założenia, których niespełnienie skutkuje obciążonymi szacunkami wielkości populacji. W przypadku jednego źródła zakładamy: (1) możliwość zidentyfikowania jednostek, (2) wielokrotną obserwację jednostek (np. dana osoba popełniła więcej niż jedno przestępstwo), (3) stałość populacji w czasie, (4) określony rozkład prawdopodobieństwa wielokrotnego wystąpienia w zbiorze danych (np. ucięty rozkład Poissona) oraz (5) niezależność kolejnych obserwacji (van der Heijden i in., 2003). Założenie o niezależności zdarzeń jest bardzo restrykcyjne i w praktyce rzadko możliwe do spełnienia (Zhang, 2008). Dlatego Godwin i Böhning (2017) zapro-

ponowali wykorzystanie dodatniego rozkładu Poissona z podwyższoną liczbą jedynek do opisu liczby wystąpień w jednym źródle, będącego wynikiem: (1) nauczenia się przez badane jednostki, jak być nierozpoznanym/uniknąć złapania lub (2) nieprzyjemności związanych z pierwszym zdarzeniem i niechęci do powtórzenia sytuacji.

W kontekście dwóch lub więcej źródeł Wolter (1986) zdefiniował następujące założenia: (1) definicje populacji we wszystkich źródłach są takie same (tj. każda jednostka z populacji ma dodatnie prawdopodobieństwo pojawienia się w wybranych źródłach), (2) populacja jest zamknięta (tj. stała w danym czasie), (3) źródła danych są niezależne, (4) brak błędów pokrycia i duplikatów, (5) brak błędów łączenia między źródłami (tj. łączenie następuje na podstawie identyfikatora) oraz (6) prawdopodobieństwo włączenia do co najmniej jednego z rejestrów powinno być jednorodne. Spełnienie tych założeń jest kluczowe w kontekście możliwości stosowania omawianych metod zarówno w przypadku dwóch, jak i wielu źródeł. Wrażliwość estymatorów wielkości populacji na złamanie powyższych założeń jest obecnie poddawana dyskusji w literaturze poświęconej statystyce publicznej (Di Cecco, Di Zio, Filippini i Rocchetti, 2018; Di Consiglio i Tuoto, 2015; Gerritse, 2016; Gerritse, van der Heijden i Bakker, 2015; Griffin, 2014; Zhang, 2015; Zhang i Dunne, 2018). Z powodu ograniczonego miejsca nie podjęto w artykule próby oceny wrażliwości na niespełnienie powyższych założeń. Planowane jest to w przyszłych pracach autorów.

W kontekście statystyki publicznej metodą capture-recapture wykorzystuje się do oceny jakości spisów w ramach badań pospisowych czy spisów kontrolnych (ang. *post-enumeration surveys* – PES albo *Census Coverage Survey* – CSS). Ocena ta polega na przeprowadzeniu niezależnego badania reprezentacyjnego w celu określenia pokrycia spisu (Gołata, 2018). Przykładowo w przypadku Narodowego Spisu Powszechnego 2002 oraz 2011 wykorzystano spisy kontrolne, jednakże ich wyniki nie zostały opublikowane (Gołata, 2012).

Metodę capture-recapture zaadaptowano także do określenia wielkości populacji wyłącznie na podstawie rejestrów administracyjnych. W takim wypadku można ją znaleźć pod pojęciem dualnej metody estymacji (ang. *dual-system estimation* – DSE), jeżeli wykorzystuje się dwa źródła danych, czy potrójnej metody estymacji (ang. *triple-system estimation* – TSE) w przypadku trzech źródeł. Na przykład Zhang i Dunne (2018) rozważali zastosowanie metody capture-recapture do estymacji populacji Irlandii na podstawie rejestru aktywności osób (Person Activity Register) będącego wynikiem łączenia 10 rejestrów administracyjnych według podejścia opartego na znakach życia (ang. *signs-of-life*) oraz ewidencji praw jazdy. Bakker, van der Heijden i Gerritse (2017) podjęli próbę estymacji liczby niezarejestrowanych rezydentów w Holandii, posługując się trzema źródłami danych: rejestrem ludności, rejestrem zatrudnionych oraz rejestrem podejrzanych o przestępstwa prowadzonym przez policję. W celu oszacowania wielkości populacji niezarejestrowanych rezydentów wykorzystano od-

powiednio zbudowany model log-liniowy, uwzględniając zmienne pomocnicze w postaci czasu pobytu, płci oraz wieku, wcześniej dokonawszy deterministycznego i probabilistycznego łączenia rekordów z trzech wspomnianych źródeł danych. Przeprowadzono również analizę wrażliwości uzyskanych wyników na przypadek występowania błędów połączenia oraz poprawności procesu parowania jednostek.

W literaturze można znaleźć również wiele innych przykładów wykorzystania omawianej metody do szacunku liczebności specyficznych subpopulacji, np. liczby bezdomnych (Coumans, Cruyff, van der Heijden, Wolf i Schmeets, 2017; Hudson, 1998; Schepers i Nicaise, 2017), narkomanów (Bouchard, 2007, 2008; Bouchard i Tremblay, 2005; van der Heijden, Cruts i Cruyff, 2013; Rossi i Mascioli, 2008), nietrzeźwych kierowców (van der Heijden i in., 2003), ofiar konfliktów (Chen, Shrivastava i Steorts, 2018) czy liczby cudzoziemców. Ciekawy przegląd zastosowań metody capture-recapture w estymacji liczebności populacji trudnych do zbadania przedstawili Godwin i Böhning (2017).

### **Metoda capture-recapture w estymacji liczby cudzoziemców**

Van der Heijden i współpracownicy (2003) rozważali wykorzystanie jednego źródła danych do estymacji liczby cudzoziemców nielegalnie przebywających w 1995 r. w Amsterdamie, Rotterdamie, Hadze oraz Utrechcie, którzy nie zostali skutecznie wydalenii z Holandii. Cudzoziemcy ci byli wielokrotnie obserwowani w zbiorach danych policji. Do estymacji wielkości tak zdefiniowanej populacji badacze zastosowali rozkład Poissona ucięty w zerze oraz odpowiadający mu uogólniony model liniowy (ang. *zero-truncated Poisson regression model*), wykorzystując następujące zmienne pomocnicze: wiek (do 40, powyżej 40 lat), płeć, narodowość (Turcja, Afryka Północna, pozostała część Afryki, Surinam, Azja, Ameryka i Australia) oraz powód wydalenia (nielegalne przebywanie, pozostałe).

Godwin i Böhning (2017) ponownie przeanalizowali zbiór danych wykorzystany przez van der Heijdena i współpracowników (2003), ale zakładając, że zdarzenia są zależne, tj. cudzoziemcy raz złapani przez policję mogą nauczyć się, w jaki sposób unikać kolejnego spotkania, lub postanowili zalegalizować swój pobyt. W tym celu zaproponowali wykorzystanie dodatkowego rozkładu Poissona z podwyższoną liczbą jedynek (pierwszych złapań) oraz uogólnionego modelu liniowego zakładającego ten rozkład dla badanej cechy. Zastosowanie tego podejścia znacząco obniżyło szacunki wielkości populacji w porównaniu z podejściem van der Heijdena i współpracowników (2003) – z 7080 do 3455. Natomiast wykorzystanie zmiennych pomocniczych zwiększyło estymowaną liczbę cudzoziemców nielegalnie przebywających na terenie wyżej wymienionych miast w 1995 r. do odpowiednio 6272 oraz 12690. Wydaje się, że w wypadku wykorzystania jednego źródła danych podejście zaproponowane przez Godwina i Böhninga (2017) jest właściwe.

W kontekście dwóch i większej liczby źródeł van der Heijden i współpracownicy (2012) przedstawili z kolei interesującą technikę estymacji osób urodzonych na Bliskim Wschodzie (Afganistan, Irak oraz Iran), ale przebywających w Holandii. W tym celu wykorzystali modele log-liniowe uwzględniające tzw. pasywne i aktywne zmienne pomocnicze. W procesie szacowania tak zdefiniowanej populacji posłużyli się dwoma rejestrami: rejestrem osób, którym wydano zezwolenie na pobyt w Holandii, oraz rejestrem policyjnym, zawierającym informacje o osobach podejrzanych o popełnienie przestępstwa.

Z kolei Gerritse i współpracownicy (2015) rozważali problem estymacji liczby Polaków oraz osób urodzonych na Bliskim Wschodzie, a przebywających w Holandii odpowiednio w 2011 i 2009 r. W tym celu sięgnęli po dane, podobnie jak van der Heijden i współpracownicy (2012), z dwóch rejestrów administracyjnych: rejestru osób zameldowanych w Holandii oraz rejestru policyjnego, skupiając się jednak na wrażliwości estymatora wielkości populacji opartego na modelach log-liniowych na złamanie założenia o niezależności tych dwóch źródeł. W przypadku osób urodzonych na Bliskim Wschodzie wpływ złamania założeń capture-recapture jest niewielki, podczas gdy dla obywateli Polski różnice w wielkości populacji są bardzo znaczące. W swojej rozprawie doktorskiej Gerritse (2016) analizowała problemy niespełnienia założeń metody capture-recapture (zależności źródeł oraz błędów w łączeniu rekordów) oraz wpływu imputacji danych na estymację liczby rezydentów według czasu przebywania. Oprócz rejestru ludności i policji autorka wykorzystwała rejestr osób zatrudnionych.

Alternatywę dla danych jednostkowych pochodzących z wielu źródeł zaproponował Zhang (2008) w kontekście estymacji subpopulacji cudzoziemców. Na potrzeby estymacji wielkości populacji odnoszącej się do nielegalnie przebywających w Norwegii cudzoziemców<sup>6</sup> sięgnął do trzech źródeł danych: Centralnego Rejestru Osób (Central Personel Register), z którego wykorzystał informacje na temat liczby zameldowanych osób urodzonych poza Norwegią według kraju urodzenia i w wieku 18 lat i więcej, zbioru danych z Krajowego Urzędu Statystycznego w Norwegii na temat liczby obcokrajowców według kraju obywatelstwa oskarżonych o popełnienie przestępstwa oraz rejestru DUF (Dataselementet for Utlendings og Flyktningsaker), w którym znajdują się wszystkie osoby ubiegające się o zamieszkanie w Norwegii (jest to baza obejmująca imigrantów i uchodźców, którym przyznawany jest 12-cyfrowy numer w momencie ubiegania się przez nich o możliwość zamieszkania w Norwegii). Z tego źródła uzyskano informację o liczbie wniosków o wydalenie z Norwegii z uwzględnieniem osób wnioskujących o azyl. Na potrzeby estymacji wielkości populacji nielegalnie przebywających w Norwegii cudzoziemców zastosowano hierarchiczny model gamma Poissona, który należy do rodziny modeli mieszanych z efektami losowymi. Jako efekt losowy wykorzystano kraj pochodzenia cudzoziemców.

---

<sup>6</sup> Autor w swojej pracy używał zamiennie pojęć *unauthorized foreigners* oraz *irregular foreigners* w kontekście rezydentów, którzy przebywali na terenie Norwegii bez wymaganych dokumentów.



Na podstawie powyższych rozważań należy zauważyć pewną powtarzalność w kontekście doboru źródeł danych. Wszystkie estymacje opierały się na rejestrze osób (populacji *de iure*) oraz danych policji. Główną przesłanką takiego wyboru jest spełnienie założenia o niezależności źródeł danych. Dlatego aby poprawnie oszacować wielkość populacji, kluczowe jest dobranie odpowiednich zbiorów administracyjnych. Od spełnienia tego założenia zależy zasadność stosowania metody capture-recapture.

Przytoczone przykłady wskazywały na praktyczne wykorzystanie metody capture-recapture bazującej na modelach log-liniowych czy Poissona w estymacji liczby cudzoziemców w innych krajach. W przypadku Polski brak jest kompleksowych opracowań skupiających się na estymacji faktycznej liczby cudzoziemców. Częściowo może wynikać to z faktu, że dopiero w ostatnich latach tematyka cudzoziemców w Polsce (zwłaszcza osób pochodzących z Ukrainy) nabrała dużego znaczenia, szczególnie w kontekście rynku pracy. Warto jednak podkreślić, że pracownicy Narodowego Banku Polskiego dokonują prób estymacji na podstawie danych zagregowanych na potrzeby modelu NECMOD (ekonometrycznego modelu polskiej gospodarki) oraz szacunków przekazów pieniężnych. Także w mediach pojawiają się różne szacunki, jednak w żaden sposób nie są weryfikowalne. Rejestr PESEL zawiera bowiem wyłącznie osoby zameldowane na pobyt czasowy lub stały, Zakład Ubezpieczeń Społecznych (ZUS) dysponuje liczbą cudzoziemców zgłoszonych do ubezpieczenia, Ministerstwo Rodziny, Pracy i Polityki Społecznej – danymi na temat oświadczeń o chęci zatrudnienia cudzoziemców, Urząd ds. Cudzoziemców (UdSC) – danymi dotyczącymi ubiegania się o wizy czy karty pobytu, straż graniczna – statystykami m.in. ruchu granicznego czy liczby cudzoziemców nielegalnie przebywających na terenie Polski, a policja – Krajowym Systemem Informacji, który zawiera dane o popełnionych przestępstwach. Mimo to wydaje się, że polska statystyka publiczna, wspierana zasobami informacyjnymi pochodzącymi od innych organów, dysponuje wszelkimi zbiorami umożliwiającymi podjęcie rzetelnej próby szacunku liczby cudzoziemców. Do tej pory, zgodnie z aktualną wiedzą autorów, w Polsce nie podejmowano prób estymacji liczby cudzoziemców z wykorzystaniem wyżej rozważanych metod. Niniejszy artykuł oraz projekt badawczy, na podstawie którego powstał, wychodzą naprzeciw oczekiwaniom wielu odbiorców w tym zakresie.

## MODELE LOG-LINIOWE W SZACOWANIU WIELKOŚCI POPULACJI TRUDNYCH DO ZBADANIA

Na potrzeby estymacji liczby cudzoziemców w Polsce z uwzględnieniem dodatkowych przekrojów zdecydowano się na zastosowanie metody capture-recapture bazującej na modelach log-liniowych. Wynikało to przede wszystkim z dostępności odpowiednich źródeł danych (które można wykorzystać w tego typu szacunkach), odpowiednich pakietów programu R (w których zaimplemen-

towane są funkcje na potrzeby estymacji parametrów modeli log-liniowych oraz kodów na procedurę bootstrap umożliwiającą znalezienie właściwych przedziałów ufności), a także z faktu, że w literaturze przedmiotu właśnie te modele są z powodzeniem wykorzystywane w estymacji liczebności populacji trudnych do zbadania. Przykładem mogą być wspomniane już prace Coumans i współpracowników (2017) oraz van der Heijdena i współpracowników (2012). W pierwszej z nich wykorzystano modele log-liniowe do oszacowania liczby bezdomnych w Holandii, w drugiej – modele log-liniowe oraz koncepcje pasywnych i aktywnych zmiennych pomocniczych do oszacowania liczby osób urodzonych na Bliskim Wschodzie, a przebywających w Holandii.

Modele log-liniowe stanowią obecnie bardzo ważną metodę analizy danych zawartych w tablicach kontyngencji. Rozwój metodologii opartej na tej technice analizy danych jakościowych zapoczątkowany został w latach 60. XX w. Goodman (1964, 1968, 1969) był jednym z pierwszych badaczy, którzy spopularyzowali modele log-liniowe w naukach społecznych. Modele te są szczególnie przydatne w sytuacji, gdy nie ma precyzyjnego rozróżnienia między zmienną objaśnianą a zmiennymi objaśniającymi, a zachodzi potrzeba wykrycia zależności w pewnym zbiorze danych.

Punktem wyjścia do zastosowania modeli log-liniowych w estymacji liczebności populacji trudnych do zbadania jest odpowiednio skonstruowana tablica kontyngencji<sup>7</sup>, w której wykorzystuje się informacje z dwóch lub większej liczby źródeł danych. Tablica 1 odnosi się do przypadku, gdy dysponujemy dwoma niezależnymi źródłami danych (A i B).

**TABL. 1. PRZYPADEK DWÓCH ŹRÓDEŁ  
– TABLICA KONTYNGENCJI 2 x 2**

A	B		Suma
	Tak (1)	Nie (0)	
Tak (1) .....	$n_{11}$	$n_{10}$	$n_{+1}$
Nie (0) .....	$n_{01}$	$n_{00}$	$n_{0+}$
Suma .....	$n_{+1}$	$n_{+0}$	$n$

U w a g a. Tak – jednostka występuje w danym źródle,  
Nie – jednostka nie występuje w danym źródle.

Ź r ó d ł o: opracowanie własne.

W wypadku dwóch źródeł danych A i B może mieć miejsce sytuacja, w której po połączeniu jednostek<sup>8</sup> występują one tylko w źródle A, a nie występują w

<sup>7</sup> Na potrzeby szacunku liczby cudzoziemców rozpatrywane były złożone tablice wielowymiarowe. W artykule ograniczono się do tablic typu 2 x 2 oraz 2 x 2 x 2 celem przedstawienia idei modeli log-liniowych w tym zagadnieniu.

<sup>8</sup> W tym celu można zastosować łączenie deterministyczne z wykorzystaniem odpowiedniego identyfikatora lub probabilistyczne łączenie rekordów.

źródle B, występują w źródle B i nie występują w źródle A lub występują jednocześnie w źródle A i B. Przykładowo w tabl. 1  $n_{01}$  oznacza liczbę jednostek, które nie występują w źródle A, a występują w źródle B. Kluczową kwestią jest zatem oszacowanie liczebności  $n_{00}$ , tj. liczby jednostek, które nie występują ani w źródle A, ani B. Ostatecznie oszacowaną liczebność populacji uzyskuje się bowiem poprzez dodanie wszystkich wartości z tabl. 1 po wcześniejszej estymacji liczebności  $n_{00}$ .

Oszacowanie liczebności  $n_{00}$  można uzyskać poprzez dopasowanie modelu log-liniowego do niekompletnej tablicy kontyngencji. Przykładowo dla tabl. 1 typu  $2 \times 2$  odnoszących się do źródeł danych A i B pełny model log-liniowy [AB]<sup>9</sup> może być przedstawiony w postaci modelu nasyconego (ang. *saturated model*):

$$\ln(m_{ij}) = \mu + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB} \quad i, j = \{\text{'Tak'}, \text{'Nie'}\} \quad (1)$$

gdzie  $m_{ij}$  oznacza oczekiwaną liczebność w komórce  $i, j$ . Ponieważ jednak komórka  $m_{00} = m_{(\text{Nie}, \text{Nie})}$  nie jest obserwowana, model [AB] ma jeden parametr za dużo i nie może być estymowany. W takiej sytuacji można rozważyć model niezależności [A][B] postaci:

$$\ln(m_{ij}) = \mu + \lambda_i^A + \lambda_j^B \quad (2)$$

w którym są tylko trzy parametry do oszacowania w związku z brakiem efektu interakcji  $\lambda_{ij}^{AB}$ . Trzy obserwowane komórki w tabl. 1 oraz trzy parametry do oszacowania w zasadzie czynią z tego modelu model nasycony. Po dopasowaniu go do danych możemy użyć oszacowanych parametrów do wyznaczenia liczebności brakującej komórki (Nie, Nie), a następnie wyznaczyć liczebność populacji poddanej analizie. Oszacowanie liczebności komórki  $n_{00}$  otrzymujemy ze wzoru:

$$\hat{n}_{00} = \exp(\mu) \quad (3)$$

Podobne rozumowanie można przeprowadzić w odniesieniu do tablic trójdzielczych typu  $2 \times 2 \times 2$ , tj. w sytuacji gdy dysponujemy trzema źródłami danych.

Tablica 2 ilustruje przypadek wykorzystania trzech źródeł (A, B, C), np. trzech rejestrów administracyjnych lub dwóch rejestrów administracyjnych i badania reprezentacyjnego czy spisu. Podobnie jak w tablicy  $2 \times 2 \times 2$  istotne jest określenie przynależności do danego źródła, a celem – oszacowanie tego, czego nie możemy odczytać z tablicy, tj.  $n_{000}$ . Na potrzeby estymacji liczebności  $n_{000}$  można również wykorzystać koncepcję modeli log-liniowych.

<sup>9</sup> Notacja nawiasowa, często stosowana w przypadku modeli log-liniowych.

**TABL. 2. PRZYPADEK TRZECH ŹRÓDEŁ – TABLICA KONTYNGENCJI 2 x 2 x 2**

A	C				Suma
	Tak (1)		Nie (0)		
	B				
	Tak (1)	Nie (0)	Tak (1)	Nie (0)	
Tak (1) .....	$n_{111}$	$n_{101}$	$n_{110}$	$n_{100}$	$n_{1++}$
Nie (0) .....	$n_{011}$	$n_{001}$	$n_{010}$	$n_{000}$	$n_{0++}$
Suma .....	$n_{+11}$	$n_{+01}$	$n_{+10}$	$n_{+00}$	$n$

U w a g a. Jak przy tabl. 1.

Ź r ó d ł o: jak przy tabl. 1.

W tym celu budujemy model log-liniowy postaci (bez efektu głównego  $\lambda_{ijk}^{ABC}$ ):

$$\ln(m_{ij}) = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} \tag{4}$$

który musimy ograniczyć przez:  $\lambda_0^A = \lambda_0^B = \lambda_0^C = \lambda_{00}^{AB} = \lambda_{10}^{AB} = \lambda_{01}^{AB} = \lambda_{00}^{AC} = \lambda_{10}^{AC} = \lambda_{01}^{AC} = \lambda_{00}^{BC} = \lambda_{10}^{BC} = \lambda_{01}^{BC} = 0$ , aby móc oszacować parametry. Dodatkowym założeniem jest to, że nie występuje interakcja między A, B i C, tj.  $\lambda_{ijk}^{ABC} = 0$ . Model ten w notacji nawiasowej oznacza się jako [AB][BC][AC]. Oszacowanie brakującej liczby jednostek populacji otrzymujemy ze wzoru:

$$\hat{n}_{000} = \exp(\mu) \tag{5}$$

po uprzednim wyestymowaniu wszystkich parametrów.

W przypadku estymacji wielkości populacji możliwe jest wykorzystanie zmiennych pomocniczych, którymi mogą być przykładowo płeć czy grupy wieku. Ma to na celu obejście jednego z założeń metody capture-recapture (o stałej stopie pokrycia przez źródło w populacji) i uwzględnienie heterogeniczności przynależności poszczególnych jednostek do źródeł. Wykorzystanie zmiennych pomocniczych w kontekście modeli log-liniowych rozważają m.in. Coumans i współpracownicy (2017), Gerritse (2016), van der Heijden i współpracownicy (2012) czy Zwane i van der Heijden (2005). Wyróżniamy przy tym dwa podejścia, które determinowane są dostępnością zmiennych we wszystkich lub niektórych źródłach, lub tylko w jednym źródle. Pierwsze określa się w literaturze jako podejście z całkowicie obserwowalnymi zmiennymi (ang. *fully observed covariates*), a drugie – z częściowo obserwowalnymi zmiennymi (ang. *partially observed covariates*). W obydwu przypadkach można wykorzystać modele log-liniowe do oszacowania poszczególnych elementów populacji. Tego typu podejście zostało również zastosowane na potrzeby tego artykułu. Przykładowo w przypadku dwuwymiarowej tablicy kontyngencji 2 x 2 oprócz przynależności do dwóch źródeł A i B można rozpatrywać dodatkową cechę X (np. płeć), przez co należy rozszerzyć

tablicę do trójdzielczej (tabl. 3) oraz dopasować model log-liniowy [AX][BX] postaci:

$$\ln(m_{ijx}) = \mu + \lambda_i^A + \lambda_j^B + \lambda_x^X + \lambda_{ix}^{AX} + \lambda_{jx}^{BX} \quad (6)$$

gdzie  $\lambda_{ix}^{AX}$  oraz  $\lambda_{jx}^{BX}$  oznaczają efekty interakcji pomiędzy zmienną pomocniczą  $X$  i źródłami danych  $A$  oraz  $B$ .

**TABL. 3. PRZYPADK DWÓCH ŹRÓDEŁ I ZMIENNEJ POMOCNICZEJ**

A	X				Suma
	$X_1$		$X_2$		
	B				
	Tak (1)	Nie (0)	Tak (1)	Nie (0)	
Tak (1) .....	$n_{111}$	$n_{101}$	$n_{110}$	$n_{100}$	$n_{1++}$
Nie (0) .....	$n_{011}$	$n_{001}$	$n_{010}$	$n_{000}$	$n_{0++}$
Suma .....	$n_{+11}$	$n_{+01}$	$n_{+10}$	$n_{+00}$	$n$

U w a g a. Jak przy tabl. 1.

Ź r ó d ł o: jak przy tabl. 1.

W przypadku dwóch źródeł  $A$  i  $B$  oraz jednej zmiennej pomocniczej  $X$ , przyjmującej przykładowo dwa warianty  $X_1$  oraz  $X_2$  (np. mężczyzna i kobieta), mamy do czynienia z trójdzielczą tablicą kontyngencji  $2 \times 2 \times 2$ , w której brakujące liczebności podlegające estymacji to  $n_{001}$  oraz  $n_{000}$ . Dla sześciu komórek zatem znane są liczebności obserwowane w tabl. 3, w związku z czym model (6) zawiera sześć parametrów, które należy oszacować (nasycony model log-liniowy). Po dopasowaniu modelu do danych brakujące liczebności komórek ustala się ze wzorów:  $\hat{n}_{000} = \exp(\mu)$  oraz  $\hat{n}_{001} = \exp(\mu + \lambda_{x_1}^X)$ . Powyższe postępowanie można rozszerzyć na większą liczbę zmiennych pomocniczych oraz źródeł. Zwiększa się przez to w oczywisty sposób złożoność analizowanych modeli log-liniowych, jednak wykorzystanie odpowiednich pakietów (np. *stats* i *parallel*) języka R (The R Development Core Team, 2018) znacznie skraca proces estymacji wszystkich możliwych do zbudowania modeli.

## METODY OCENY JAKOŚCI MODELI LOG-LINIOWYCH

W analizie log-liniowej głównym celem jest wybór modelu o możliwie najprostszej postaci, który jednocześnie byłby najlepiej dopasowany do danych. W literaturze przedmiotu (Brzezińska, 2015; Goodman, 1964, 1968, 1969) proponuje się różnego rodzaju kryteria oceny modeli. Zostały one również wykorzystane na potrzeby niniejszego artykułu w procesie wyboru i oceny finalnego modelu. Do najważniejszych kryteriów zaliczamy iloraz wiarygodności, dewiancję,  $AIC$  (Akaike Information Criterion) oraz  $BIC$  (Bayerian Information Criterion).

Iloraz wiarygodności jest miarą pozwalającą ocenić dopasowanie modelu do danych. Przykładowo dla tablic 2 x 2 wyraża się on wzorem:

$$G^2 = 2 \sum_{i=1}^2 \sum_{j=1}^2 n_{ij} \ln \left( \frac{n_{ij}}{\hat{m}_{ij}} \right) \quad (7)$$

gdzie  $\hat{m}_{ij} = \frac{n_{i+}n_{+j}}{n}$  stanowią oszacowania liczebności teoretycznych wyznaczonych dla danego modelu log-liniowego. W sytuacji gdy wartość ilorazu wiarygodności  $G^2$  jest duża, model taki powinien być odrzucony jako ten, który w nieprawidłowy sposób odwzorowuje zależności między badanymi zmiennymi. Współczynnik  $G^2$  może być także wykorzystywany do porównania oceny różnych modeli. Jeżeli porównuje się dwa modele, współczynnik  $G^2$  może zostać przedstawiony w postaci (dla tablic 2 x 2):

$$G^2 = 2 \sum_{i=1}^2 \sum_{j=1}^2 \hat{m}_{ij}^0 \ln \left( \frac{\hat{m}_{ij}^0}{\hat{m}_{ij}^1} \right) \quad (8)$$

gdzie 0 odnosi się do liczebności teoretycznych modelu ogólniejszego, tj. zawierającego wszystkie możliwe parametry, natomiast 1 dotyczy liczebności teoretycznych modelu zagnieżdżonego o uproszczonej postaci i zawierającego się w modelu 0.

Współczynnik ten może być również przedstawiony w postaci:

$$G^2(M_0|M_1) = G^2(M_0) - G^2(M_1) \quad (9)$$

Powyższa statystyka ma rozkład chi-kwadrat o liczbie stopni swobody  $df = df(M_0) - df(M_1)$ , gdzie  $M_0$  jest modelem zagnieżdżonym, a  $M_1$  modelem ogólnym z większą liczbą parametrów, i nazywana jest dewiancją. Dewiancja pozwala ocenić, czy parametr występujący w modelu  $M_1$ , a niewystępujący w modelu  $M_0$  jest statystycznie istotny.

Statystyką służącą do porównywania ze sobą większej liczby modeli jest kryterium informacyjne Akaike oraz Schwarza (bayesowskie). Kryterium informacyjne Akaike wyraża się wzorem:

$$AIC = G^2 - df \quad (10)$$

gdzie  $G^2$  to iloraz wiarygodności badanego modelu, a  $df$  to liczba odpowiadających mu stopni swobody. Z kolei bayesowskie kryterium informacyjne wyraża się wzorem:

$$BIC = G^2 - df \cdot \ln(n) \quad (11)$$

gdzie  $n$  to liczebność w tablicy kontyngencji. Preferowane są przy tym modele, dla których miary  $AIC$  i  $BIC$  przyjmują mniejsze wartości. W pracy wykorzystano kryterium  $BIC$  do określenia najlepszego modelu.

## PRECYZJA OSZACOWAŃ LICZEBNOŚCI POPULACJI TRUDNEJ DO ZBADANIA

Kluczową kwestią w zagadnieniu estymacji liczebności populacji trudnej do zbadania jest jakość uzyskanych wyników. Oceną precyzji oszacowań uzyskanych za pomocą technik capture-recapture zajmowali się liczni badacze oraz wiele instytucji. Przykładowo Międzynarodowa Grupa Robocza ds. Monitorowania i Prognozowania Chorób (International Working Group for Disease Monitoring and Forecasting) prowadzi prace nad konstrukcją niesymetrycznych przedziałów ufności dla liczebności populacji trudnej do zbadania. Z kolei Chao (1989) podjął próbę konstrukcji symetrycznych przedziałów ufności polegającą na odpowiedniej transformacji oszacowanej liczebności populacji, głównie z wykorzystaniem transformacji logarytmicznej. Wreszcie ostatnio stosowane techniki w konstrukcji estymatorów wariancji liczebności populacji trudnej do zbadania bazują na metodzie bootstrap, zarówno nieparametrycznej, jak i parametrycznej (Buckland i Garthwaite, 1991; Gemmell, Millar i Hay, 2004).

W artykule na potrzeby oceny jakości oszacowań liczby cudzoziemców w odpowiednich przekrojach skonstruowano 95-procentowe przedziały ufności oraz względne błędy szacunku. W tym celu wykorzystano parametryczny bootstrap, który jest szeroko stosowany w badaniach poświęconych estymacji populacji trudnych do zbadania (Zwane i van der Heijden, 2003). Decyzja o konstrukcji odpowiednich przedziałów ufności oraz względnych błędów szacunku bazujących na parametrycznej metodzie bootstrap wynikała również z tego, że jest to stosunkowo łatwa w implementacji technika w kontekście tablic kontyngencji, które nie są w pełni obserwowalne (nieznajomość liczebności niektórych komórek). Ogólnie w celu utworzenia przedziałów ufności oraz wyznaczenia względnych błędów szacunku w pierwszej kolejności dokonuje się oszacowania liczebności populacji trudnej do zbadania z wykorzystaniem odpowiedniego modelu log-liniowego. Estymację parametrów modelu log-liniowego przeprowadza się na obserwowalnych komórkach tablicy kontyngencji. Na podstawie oszacowanych parametrów modelu oraz liczebności brakujących komórek można wyznaczyć prawdopodobieństwo teoretyczne przynależności dla wszystkich komórek w tablicy kontyngencji. Następnie losuje się próbę z rozkładu wielomianowego przy uwzględnieniu oszacowanych prawdopodobieństw, która na dalszym etapie jest korygowana, tak aby odpowiadała strukturze obserwowanych danych. Wówczas dopasowuje się odpowiedni model log-liniowy do kompletnej tablicy kontyngencji i uzyskuje pierwsze oszacowanie bootstrapowe liczebności populacji trudnej do zbadania. Procedurę tę przeprowadza się wielokrotnie, wyznaczając wariancję, a następnie przedział ufności dla liczebności populacji.

Ujmując zagadnienie bardziej formalnie, sposób wyznaczania względnych błędów szacunku oraz przedziałów ufności liczebności populacji trudnej do zbadania w parametrycznej metodzie bootstrap można zapisać w następujących krokach:

1. Oszacowanie parametrów odpowiedniego modelu log-liniowego dla zadanej tablicy kontyngencji i komórek, dla których istnieją wartości empiryczne.
2. Oszacowanie wielkości populacji we wszystkich założonych przekrojach z wykorzystaniem parametrów wyznaczonego modelu log-liniowego.
3. Wyznaczenie całkowitej liczebności populacji trudnej do zbadania  $\hat{N} = \hat{N}_1 + \dots + \hat{N}_p$ , gdzie  $P$  to liczba komórek w tablicy kontyngencji, a  $\hat{N}_p$  to oszacowana wielkość populacji w komórce  $p$ , przy czym  $p = 1, \dots, P$ .
4. Wyznaczenie wektora długości  $P$  złożonego z prawdopodobieństw  $\hat{\pi}_p = (\hat{N}_1/\hat{N}, \dots, \hat{N}_p/\hat{N})^T$ .
5. Generowanie z rozkładu wielomianowego wektora  $\mathbf{n}^* = (N_1^*, \dots, N_p^*)^T$  długości  $P$  odpowiadającego populacji o liczebności  $\hat{N}$  z prawdopodobieństwami  $\hat{\pi}_p$ . Jest to wektor złożony z pseudoliczebności populacji we wszystkich przyjętych  $P$  przekrojach.
6. Utworzenie tablicy kontyngencji na bazie uzyskanej pseudoliczebności, układem odpowiadającej tablicy wyjściowej. Oszacowane są parametry modelu log-liniowego dla tych samych komórek co w punkcie 1.
7. Znajdowanie oszacowania liczebności dla przekrojów nieobserwowanych w tablicy kontyngencji.
8. Powtórzenie  $B$  razy<sup>10</sup> kroków 5–7.
9. Oszacowanie liczebności populacji  $\hat{N}^b$ , dla  $b = 1, \dots, B$  oraz liczebności we wszystkich przekrojach rozważanej tablicy kontyngencji.
10. Wyznaczenie na podstawie otrzymanych oszacowań wartości oczekiwanej, wariancji, względnego błędu szacunku oraz 95-procentowego przedziału ufności liczebności populacji trudnej do zbadania<sup>11</sup>:
  - wartość oczekiwana:

$$\hat{N} = \frac{\sum_{b=1}^B \hat{N}^b}{B} \quad (12)$$

- wariancja empiryczna:

$$\hat{V}(\hat{N}) = \frac{1}{B-1} \sum_{b=1}^B (\hat{N}^b - \hat{N})^2 \quad (13)$$

<sup>10</sup> Na potrzeby artykułu przyjęto  $B = 500$ .

<sup>11</sup> W podobny sposób można wyznaczyć te miary dla liczebności populacji w odpowiednich przekrojach.



- względny błąd szacunku (precyzja):

$$REE(\hat{N}) = \frac{\sqrt{\hat{V}(\hat{N})}}{\hat{N}} \quad (14)$$

- 95-procentowy przedział ufności<sup>12</sup>:

$$[\hat{N}_{2,5\%}, \hat{N}_{97,5\%}]$$

## ŹRÓDŁA DANYCH

### Wybór zbiorów

Postawione w artykule cele realizowane były z wykorzystaniem danych pochodzących z zasobów informacyjnych statystyki publicznej za lata 2015 i 2016 (PBSSP), w szczególności danych administracyjnych i statystycznych gromadzonych w ramach badań: Zasoby migracyjne w Polsce, Cudzoziemcy w Polsce. Legalizacja pobytu cudzoziemców na terytorium RP, Operat do Badań Społecznych, Charakterystyka demograficzno-społeczna i ekonomiczna gospodarstw domowych i rodzin oraz Badanie Aktywności Ekonomicznej Ludności.

Po dokonaniu analizy i oceny zasobów danych jako główne źródła administracyjne w modelach log-liniowych posłużyły:

- System Pobyt (UdSC) – zbiór rejestrów, ewidencji i wykazu w sprawach cudzoziemców w zakresie wydanych zezwoleń na pobyt;
- rejestr PESEL (Ministerstwo Cyfryzacji) – w przypadku cudzoziemców zameldowanych wyłącznie na pobyt stały;
- Centralny Rejestr Ubezpieczonych (ZUS) – w przypadku ubezpieczonych cudzoziemców oraz członków ich rodzin (udostępniony zbiór nie obejmował wszystkich ubezpieczonych)<sup>13</sup>.

### Przygotowanie zbiorów danych do badania

Zbiory wejściowe wykorzystane w projekcie badawczym poddano przetwarzaniu umożliwiającemu porównywanie, łączenie i analizę danych z różnych źródeł oraz oszacowanie wyników. Można wyróżnić kilka, wzajemnie się przenikających i uzupełniających, grup działań:

<sup>12</sup> Przedział ten wyznaczany jest metodą percentylową, np. 95-procentowy percentylowy przedział ufności ma dolną i górną granicę wyznaczoną przez 2,5 i 97,5 percentyla wartości bootstrapowych  $\hat{N}^b$ .

<sup>13</sup> W projekcie, który jest podstawą tego artykułu, wykorzystano więcej źródeł, nieuwzględnionych tutaj z racji innego zastosowania oraz ograniczonego miejsca.

1. Dobór podmiotowy i przedmiotowy. W tej fazie prac – na podstawie wstępnej analizy zawartości zbiorów wejściowych oraz stosownie do przyjętego zakresu przedmiotowego badania i przesłanek metodologicznych – dokonano selekcji potencjalnie przydatnych zmiennych ze zbiorów. Z kolei stosownie do zakresu podmiotowego badania zastosowano dobór rekordów w taki sposób, aby dotyczyły one cudzoziemców (np. w przypadku zbiorów z badań obejmujących szersze kategorie ludności) w odpowiednich do przyjętych w badaniu momentów obserwacji (31 grudnia 2015 r. i 2016 r.) pod względem okresu przebywania w Polsce (w przypadku zbiorów rejestrowych odnotowujących fakty i daty dotyczące pobytu).
2. Wyliczanie cech pochodnych na podstawie przekształceń surowych danych. W ramach tej grupy działań wykonano szereg wyliczeń i przekształceń surowych danych, mających na celu przede wszystkim: (1) utworzenie (wyprowadzenie) cech potrzebnych do opisu badanej populacji, czyli np. wyliczenie okresu pobytu cudzoziemca na podstawie dat zarejestrowanych w dokumentach; (2) zapewnienie zgodności definicyjnej i zakresowej cech pochodzących z różnych źródeł, np. dostosowanie różnorodnych konwencji zapisów kraju obywatelstwa do ujednoliconego słownika kodów krajów.
3. Redukcja nadmiarowych danych. Ta faza składała się z dwóch kroków: (1) deduplikacji w obrębie pojedynczych zbiorów danych, polegającej na wykrywaniu i usuwaniu ewidentnych dubli – powielonych rekordów danych, czyli takich, które pomimo różnych technicznych (bazodanowych) identyfikatorów rekordu zawierały dokładne powtórzenie wszystkich wartości; (2) niwelowania redundancji podmiotowej danych w kilku zbiorach jednego rejestru wartości/danych. Dokonano łączenia (parowania) poszczególnych zbiorów w ramach danego rejestru i wykrywania rekordów dotyczących tych samych podmiotów (osób), a następnie – na podstawie przesłanek merytorycznych i utworzonych na ich podstawie hierarchii adekwatności – wyboru najodpowiedniejszego rekordu reprezentującego danego cudzoziemca w rejestrze. W konsekwencji w odniesieniu do określonego rejestru powstawał – w zależności od potrzeb – jeden zbiór zawierający dane dotyczące unikalnych jednostek lub kilka zbiorów, ale podmiotowo rozłącznych.
4. Wyodrębnianie podstawowych jednostek/podmiotów badania. W wielu zbiorach rejestrowych obejmujących cudzoziemców podstawowe jednostki danych (rekordy) nie odnoszą się bezpośrednio do pojedynczych osób, lecz do różnego rodzaju faktów ich dotyczących. Stąd niezbędne były przekształcenia zbiorów wejściowych, w wyniku czego otrzymywano rekordy danych odnoszące się do osób. W szczególności były to działania oparte na: (1) grupowaniu (agregowaniu) rekordów danych, w ramach którego utworzono rekordy dla osób oraz wyprowadzono za pomocą operacji i funkcji agregujących przewidziane w badaniu cechy charakteryzujące cudzoziemców lub cechy pomocnicze; (2) restrukturyzacji (transpozycji) danych, w wyniku których pewne różno-

rodne wartości dotyczące jednej osoby zarejestrowane w kilku rekordach (różne warianty cechy) zapisywano w kolumnach jednego rekordu odnoszącego się do osoby.

5. Łączenie (parowanie) rekordów z różnych zbiorów. Operacje łączenia przeprowadzane były zarówno w obrębie zbiorów pochodzących z jednego rejestru – zazwyczaj na podstawie przygotowanego przez gestora sztucznego identyfikatora rekordów/osób – jak i kojarzenia zbiorów z różnych rejestrów czy badań, w tym wypadku na ogół za pomocą uniwersalnego identyfikatora (numer PESEL) lub na podstawie kombinacji wartości kilku cech<sup>14</sup>.

## WYNIKI BADANIA

### Spełnianie założeń metody capture-recapture

W związku z wykorzystaniem w badaniu metody capture-recapture bazującej na wielu źródłach, w celu oszacowania liczby cudzoziemców poza dostępnymi statystycznymi źródłami danych, w pierwszej kolejności określono założenia metodologiczne. Kluczowe założenia, których spełnienie jest niezbędne z punktu widzenia przyjętych rozwiązań modelowych, oraz podjęte działania są następujące:

1. Definicje populacji we wszystkich rozważanych źródłach są takie same – określono populację cudzoziemców jako osoby w wieku 18 lat i więcej posiadające obywatelstwo inne niż polskie, które przebywały w Polsce pod koniec 2015 i 2016 r. Każde z wykorzystanych źródeł zostało ograniczone do tej populacji.
2. Populacja jest zamknięta – zakłada się, że w badanym okresie wielkość populacji jest stała. Ponadto należy podkreślić, że wszystkie rejestry były aktualne na ten sam dzień, tj. 31 grudnia 2016 r., dlatego podjęto następujące kroki przy wyodrębnianiu populacji:
  - populacja na 31 grudnia 2015 r.:
    - na podstawie rejestrów PESEL i ZUS oraz zbioru UdSC wybrano tylko osoby urodzone przed 31 grudnia 1997 r.,
    - na podstawie zbioru UdSC wybrano tylko te osoby, które miały decyzję umożliwiającą pobyt w Polsce wydaną między 1 stycznia a 31 grudnia 2015 r.;

---

<sup>14</sup> W łączeniu zbiorów wykorzystanych w niniejszym artykule zastosowano również parowanie według kluczy alternatywnych wobec numeru PESEL – głównie w odniesieniu do zbioru UdSC, w którym znaczna część rekordów nie miała numeru PESEL. Wykorzystano w nich, jako klucz podstawowy, zestawienia uwzględniające datę urodzenia, płeć i kraj obywatelstwa oraz – w zależności od rodzaju podejścia i dostępności zapisów w kolumnach – różne kombinacje spośród takich cech, jak: kod gminy, nazwa miejscowości i numer budynku. W tym wypadku liczba połączeń niejednoznacznych była stosunkowo niewielka i ostatecznie zrezygnowano z łączenia stochastycznego.

- populacja na 31 grudnia 2016 r.:
  - na podstawie rejestrów PESEL i ZUS oraz zbioru UdSC wybrano tylko osoby urodzone przed 31 grudnia 1998 r.,
  - na podstawie zbioru UdSC wybrano tylko te osoby, które miały decyzję umożliwiającą pobyt w Polsce wydaną między 1 stycznia a 31 grudnia 2016 r.

W wypadku zbioru UdSC nie wyłączono z analizy cudzoziemców, którym data ważności wydanego dokumentu upłynęła w ciągu 2016 r., tj. przed 31 grudnia 2016 r., ponieważ mogli przebywać w Polsce nielegalnie.

3. Źródła danych są niezależne – w przypadku źródeł administracyjnych systemy powinny być niezależne (w sensie statystycznym), aby możliwe było zastosowanie metody capture-recapture wykorzystującej modele log-liniowe. Niezależność w kontekście źródeł administracyjnych oznacza, że prawdopodobieństwo znalezienia się jednostki w jednym źródle nie zależy od przynależności tej jednostki do drugiego źródła. Ostatecznie na potrzeby artykułu wykorzystano kombinację trzech źródeł danych, które umożliwiają spełnienie tego założenia (tj. PESEL, UdSC i ZUS). Głównym uzasadnieniem wyboru tych źródeł danych było ich bieżące wykorzystywanie w statystyce publicznej na potrzeby innych badań (nie wymagało to pozyskania danych spoza PBSSP) oraz objęcie tej samej populacji.
4. Brak błędów nadreprezentacji i duplikatów – zakłada się, że źródła są pozbawione błędów nadreprezentacji, tj. zawierają wyłącznie jednostki z badanej populacji oraz zostały zdeduplikowane. Podstawowym źródłem był zintegrowany zbiór danych powstały w wyniku łączenia kilku rejestrów administracyjnych i zdeduplikowany. Zawierał on zmienną dotyczącą jakości danego rekordu, który jest przybliżeniem błędu nadreprezentacji. Z rekordów występujących w rejestrach PESEL i ZUS lub zbiorze UdSC wyodrębniono te, dla których określono kody jakości: 1 – oznaczający sytuację referencyjną (potwierdzenie istnienia osoby), 3 – wskazujący osoby w wieku 90 lat i więcej oraz 6 – oznaczający osobę zidentyfikowaną tylko w jednym rejestrze, który był wyznaczony przed dołączeniem zbioru UdSC. Dodatkowo przyjęto przy tym założenie, że cudzoziemcy będący w rejestrach przebywają na terenie Polski, niezależnie od tego, czy mają ustalone miejsce pobytu. Jest to kluczowe zwłaszcza w przypadku Systemu Pobyt, którego gestorem jest UdSC.
5. Każdą jednostkę będzie można zidentyfikować i połączyć między źródłami bez błędów – w tym celu zintegrowano dane za pomocą identyfikatora PESEL lub kombinacji zmiennych jednoznacznie wskazujących daną osobę (łączenie deterministyczne). Nie dokonywano łączenia probabilistycznego.
6. Prawdopodobieństwo włączenia do co najmniej jednego z rejestrów powinno być jednorodne – aby spełnić to założenie, w procesie estymacji wykorzystano modele zawierające następujące zmienne: kraj obywatelstwa, płeć, wiek

(2 grupy) i województwo (16 oraz nieustalone). Wybór zmiennych podyktowany był z jednej strony ich dostępnością, z drugiej zaś koniecznością spełnienia warunku, aby w odpowiednio utworzonych grupach prawdopodobieństwo włączenia cudzoziemca do danego źródła było jednakowe. Jest to jeden ze sposobów spełnienia założenia dotyczącego homogeniczności prawdopodobieństw, który rekomenduje się w literaturze poświęconej metodzie capture-recapture (van der Heijden i in., 2012).

### Opis danych

W tabl. 4 przedstawiono liczbę cudzoziemców według występowania w trzech źródłach dla analizowanych lat.

**TABL. 4. LICZBA CUDZOZIEMCÓW W WIEKU 18 LAT I WIĘCEJ**  
(stan na 31 grudnia)

Wyszczególnienie			UdSC		Suma	
			Nie	Tak		
<b>2015</b>						
PESEL	Nie	ZUS	Nie	x	30090	30090
			Tak	3821	5583	9404
	Tak		Nie	7042	7476	14518
			Tak	4620	9871	14491
Suma			15483	53020	68503	
<b>2016</b>						
PESEL	Nie	ZUS	Nie	x	92106	92106
			Tak	3821	11224	15045
	Tak		Nie	7115	16549	23664
			Tak	4641	18951	23592
Suma			15577	138830	154407	

U w a g a. Jak przy tabl. 1, x – nieznana liczba cudzoziemców poza rejestrami.

Ź r ó d ł o: opracowanie własne na podstawie rejestrów PESEL i ZUS oraz zbioru UdSC.

W przypadku kombinacji PESEL, UdSC i ZUS w 2015 r. wykorzystano informacje o ponad 68,5 tys., a w 2016 r. – blisko 154 tys. cudzoziemców.

W odniesieniu do 2015 r. jedynie 9871 cudzoziemców było zidentyfikowanych jednocześnie w rejestrach PESEL i ZUS i zbiorze UdSC, a w 2016 r. – 18951. Głównym celem jest zatem oszacowanie liczby cudzoziemców będących poza tymi rejestrami, tj. nieznannej wartości liczbowej na przecięciu pól: PESEL = Nie, UdSC = Nie i ZUS = Nie. Na potrzeby estymacji tej liczebności wykorzystano modele log-liniowe.

## Dobór modelu

Zgodnie z literaturą poświęconą szacowaniu wielkości nieznanej populacji założono, że prawdopodobieństwo pokrycia przez określone źródła danych nie jest jednakowe. Dlatego na potrzeby procesu modelowania wykorzystano następujące zmienne:

- płeć – 1 = mężczyzna, 2 = kobieta;
- wiek – produkcyjny (18–59 lat dla kobiet, 18–64 lat dla mężczyzn), poprodukcyjny – 60 lat i więcej dla kobiet, 65 lat i więcej dla mężczyzn;
- kraj obywatelstwa – UE, Armenia, Mołdawia, Białoruś, Rosja, Ukraina, Wietnam, pozostałe;
- województwo: 16 województw kodowanych 1, ..., 16, nieustalone (jeżeli nie zostało określone miejsce pobytu).

Na potrzeby wyboru końcowego modelu log-liniowego, który wykorzystano w procesie estymacji liczby cudzoziemców w Polsce w odpowiednich przekrojach, w pierwszej kolejności dokonano zakodowania zmiennych, zgodnie z symbolicznym zapisem (notacja nawiasowa) charakterystycznym dla modeli log-liniowych.

Procedurę modelowania przeprowadzono oddzielnie dla lat 2015 i 2016 oraz dla kombinacji źródeł. Oznacza to, że ostatecznie przeprowadzono dwie niezależne procedury szacunku wielkości populacji (tabl. 5).

**TABL. 5. WYBRANE MIARY JAKOŚCI MODELI LOG-LINIOWYCH WEDŁUG ROKU**

Modele	Dewiancja $M_0$	$df M_0$	$G^2$	$AIC$	$BIC$	Dewiancja	$df r$
<b>2015</b>							
1 .....	216672	2218	-24972	50003	50168	41580	2190
2 .....	216672	2218	-8909	18349	19860	9453	1954
2s .....	216672	2218	-8909	18349	19860	9453	1954
<b>2016</b>							
1 .....	651403	2399	-50619	101295	101463	91543	2371
2 .....	651403	2399	-12260	25051	26583	14827	2135
2s .....	651403	2399	-12260	25049	26575	14827	2136

U w a g a. 1 – model wyłącznie z efektami głównymi, 2 – model z efektami głównymi i interakcjami pierwszego rzędu, 2s – model 2 z zastosowaną procedurą krokową (s pochodzi od ang. *step*, które odnosi się do pojęcia regresji krokowej, ang. *stepwise selection*; *stepwise regression*),  $df$  – stopnie swobody,  $M_0$  – model jedynie z wyrazem wolnym (pusty model),  $df r$  – różnica między liczbą stopni swobody modelu pustego a liczbą stopni swobody modelu w danym wierszu.

Ź r ó d ł o: jak przy tabl. 4.

Tablica 5 zawiera zestawienie wybranych miar jakości dla zastosowanych modeli log-liniowych według kombinacji źródeł oraz roku. W przypadku modelu opartego na kombinacji PESEL, UdSC i ZUS modele 2 i 2s okazały się iden-

tyczne w 2015 r. (co potwierdzają kryteria informacyjne), a w przypadku 2016 r. model 2s był nieznacznie lepszy od modelu 2, ponieważ zarówno kryteria informacyjne *AIC*, jak i *BIC* są niższe.

### Estymacja punktowa i przedziałowa

Tablica 6 przedstawia finalny model wraz z oszacowaną wielkością populacji cudzoziemców w Polsce w latach 2015 i 2016 oraz 95-procentowym bootstradowym przedziałem ufności. Model dla 2015 r. różni się od modelu dla 2016 r. dodatkowym elementem – interakcją między płcią a wiekiem. Wynik modelowania sugeruje, że prawdopodobieństwo pokrycia przez badane źródła danych jest stałe, co skutkuje stabilnością modelu w czasie.

Według szacunków liczba cudzoziemców w wieku 18 lat i więcej przebywających w Polsce na koniec 2015 r. wynosiła 507,7 tys. (95-procentowy przedział ufności – od 369,1 tys. do 724,4 tys.). Liczba ta – oprócz cudzoziemców zameldowanych na pobyt czasowy – obejmowała również cudzoziemców zameldowanych na pobyt stały (takich osób według rejestru PESEL było 39,1 tys.). Dla porównania, zgodnie z danymi ZUS, liczba ubezpieczonych cudzoziemców zgłoszonych do ubezpieczeń emerytalnych i rentowych wynosiła 184188 na koniec 2015 r.

TABL. 6. OSTATECZNY MODEL LOG-LINIOWY

Model	$\hat{N}$	Przedział ufności	Precyzja w %
<b>2015</b>			
[P][Z][U][V][S][A][C][PZ][PU][PV][PS][PA][PC] [ZU][ZV][ZA][ZC][UV][UC][VS][VA][VC][SA][AC] [UA][US][ZS][SC] .....	507693	(369135, 724407)	17,64
<b>2016</b>			
[P][Z][U][V][S][A][C][PZ][PU][PV][PS][PA][PC] [ZU][ZV][ZA][ZC][UV][UC][VS][VA][VC][SA][AC] [UA][US][SC] .....	743665	(600796, 943124)	11,70

U w a g a. P – PESEL, U – UdSC, Z – ZUS, S – płeć, A – wiek, SA – interakcja między płcią a wiekiem, C – kraj obywatelstwa, V – województwo. Notacja nawiasowa oznacza efekty główne oraz interakcje.

Ź r ó d ł o: jak przy tabl. 4.

Analogicznie oszacowano liczbę cudzoziemców w wieku 18 lat i więcej przebywających w Polsce pod koniec 2016 r. na 743,7 tys. (95-procentowy poziom ufności – 600,8–943,1 tys.). Liczba ta oprócz cudzoziemców zameldowanych na pobyt czasowy obejmowała również cudzoziemców zameldowanych na pobyt

stały (takich osób według rejestru PESEL było 41,4 tys.). W 2016 r. odnotowano wyraźny wzrost liczby cudzoziemców w stosunku do roku poprzedniego. Zwiększyła się liczba obywateli Ukrainy, Białorusi, Rosji, Wietnamu i innych krajów spoza UE, natomiast liczba obywateli UE nieznacznie spadła. Według statystyk ZUS liczba osób fizycznych mających inne obywatelstwo niż polskie zgłoszonych do ubezpieczeń społecznych i rentowych pod koniec 2016 r. wynosiła 293188, natomiast liczba pracowników cudzoziemców zgłoszonych do tego samego ubezpieczenia – 169350. Tablica 7 przedstawia szczegółowe zestawienie wyników w podziale na kraj obywatelstwa.

**TABL. 7. SZACUNEK WIELKOŚCI POPULACJI CUDZOZIEMCÓW W POLSCE WEDŁUG KRAJU OBYWATELSTWA**

Kraje	N	Przedział ufności 95%		Precyzja w %	
		dolna granica	górna granica		
Armenia .....	2015	3168	2263	4505	18,33
	2016	4773	3897	6032	11,35
Białoruś .....	2015	19868	14429	27951	17,38
	2016	25813	20832	32569	11,81
Mołdawia .....	2015	2693	1613	4227	25,59
	2016	7580	5355	10617	17,99
Rosja .....	2015	22611	16040	32237	18,62
	2016	25534	20685	32344	12,07
Ukraina .....	2015	283714	203946	415732	18,55
	2016	454974	361512	584696	12,27
Wietnam .....	2015	7408	5554	9942	15,45
	2016	11728	10008	14170	9,10
UE .....	2015	70901	53579	97126	15,63
	2016	59571	50914	71169	8,77
Pozostałe .....	2015	97329	70037	138339	17,86
	2016	153692	124170	196140	12,06

Źródło: jak przy tabl. 4.

Wśród cudzoziemców przebywających w Polsce zdecydowanie przeważają obywatele krajów trzecich (co oznacza każdą osobę, która nie jest obywatelem UE w rozumieniu art. 17 ust. 1 Traktatu o Unii Europejskiej, w tym bezpaństwowców). Polski rynek pracy jest atrakcyjny dla cudzoziemców zza wschodniej granicy ze względu na bliskość geograficzną, sieci migracyjne, które pozwalają zminimalizować koszty pobytu przynajmniej w pierwszych tygodniach, zdecydowanie wyższe zarobki niż w krajach rodzimych, a także z powodu liberalizacji zasad dostępu obywateli do polskiego rynku pracy. Uregulowania prawne wprowadzające uproszczoną procedurę zezwalają na podejmowanie pracy (oświadczenia pracodawców o powierzeniu pracy cudzoziemcowi) przez obywateli sześciu krajów trzecich: Armenii, Białorusi, Gruzji, Mołdawii, Rosji i Ukrainy. Spośród nich obywatele Ukrainy stanowią największą zbiorowość. Szacuje się, że w 2015 r. w Polsce przebywało 283,7 tys. (95-procentowy przedział ufności –



od 203,9 tys. do 415,7 tys.), a w 2016 r. – 455,0 tys. (95-procentowy przedział ufności – od 361,5 tys. do 584,7 tys.) obywateli tego kraju.

## PODSUMOWANIE

Wybrana do oszacowania liczby cudzoziemców na krajowym rynku pracy metoda capture-recapture, bazująca na modelach log-liniowych, jak dotąd nie była stosowana w badaniach statystycznych w Polsce. Jedynymi doświadczeniami, z których można było skorzystać, są empiryczne badania zrealizowane przez badaczy holenderskich i norweskich.

Przedstawione w artykule wyniki estymacji wielkości populacji cudzoziemców, będące pochodną rezultatów otrzymanych w projekcie badawczym, mogą stanowić dobrą podstawę do wyprowadzenia (wtórnie) różnego rodzaju wskaźników dla wyodrębnionych jednostek terytorialnych, takich jak np. bilans migracyjny netto czy wskaźnik aktywności zawodowej cudzoziemców. Te ostatnie zaś mogą być wykorzystywane przez władze samorządowe, m.in. do monitorowania wielkości zatrudnienia, wysokości stopy bezrobocia czy popytu na pracę cudzoziemców o wysokich kwalifikacjach oraz do oceny wpływu powierzania pracy cudzoziemcom na wysokość płac pracowników rodzimych. Dodatkowo przedstawione w artykule szacunki mogą zostać użyte do monitorowania grup narażonych na wykluczenie zawodowe i społeczne poprzez zapobieganie substytucji rodzimych zasobów pracy przez cudzoziemców. Ponadto będą mogły stanowić podstawę do prowadzenia analiz statystycznych dotyczących sytuacji społeczno-gospodarczej poszczególnych regionów kraju oraz prognoz ich rozwoju.

Przedstawione wyniki mają innowacyjny charakter ze względu na zastosowaną metodę opracowania szacunku oraz wykorzystane źródła danych. Należy jednak zaznaczyć, że w trakcie badań napotkano poważne trudności. Wynikały one głównie z tego, że metoda szacunku opierała się na źródłach danych administracyjnych pozyskiwanych w ramach PBSSP do innych badań, a zatem o zakresie informacyjnym zdefiniowanym przez określoną jednostkę realizującą własne badanie. Jako przykład można wskazać rejestry ZUS pozyskiwane na potrzeby realizacji badań z zakresu rynku pracy, które nie zawierały informacji o wszystkich ubezpieczonych cudzoziemcach. Bardzo cenne zbiory dotyczące zezwoleń na pracę i oświadczeń pracodawców o zamiarze powierzenia pracy cudzoziemcowi nie uwzględniały cech identyfikacyjnych, w związku z czym nie można było ich połączyć deterministycznie z innymi zbiorami. Rejestr PESEL z kolei nie obejmował cudzoziemców przebywających czasowo i zameldowanych w gminach, którzy nie posiadali numeru PESEL. Co więcej, wszystkie wykorzystane rejestry były aktualne na 31 grudnia 2016 r., co mogło wpłynąć na wyniki uzyskane na 31 grudnia 2015 r. Tym samym wtórne wykorzystanie rejestrów i zawartych w nich zmiennych miało istotny wpływ zarówno na sam wybór źródeł danych oraz metodę, jak i w konsekwencji na konstrukcję wskaźników.

Niezbędne jest podjęcie prac nad rozpoznaniem również innych źródeł, które ze względu na swój charakter i zakres mogą być bardzo przydatne do weryfikacji charakterystyki cudzoziemców, np. rejestr policji dotyczący cudzoziemców podejrzanych o popełnienie przestępstw czy zbiory: Państwowej Inspekcji Pracy w zakresie kontroli legalności zatrudniania cudzoziemców, Komendy Głównej Straży Granicznej w zakresie legalności pobytów lub Ministerstwa Spraw Zagranicznych w zakresie wiz. W tym celu konieczne będzie nawiązanie bądź zintensyfikowanie współpracy z gestorami poszczególnych rejestrów i baz danych. Jednocześnie należy podkreślić, że w dalszych pracach planuje się wykorzystanie zarówno metod uwzględniających łączenie deterministyczne i probabilistyczne oraz analizy wrażliwości na złamanie założeń metody capture-recapture, jak i stosowanych modeli.

#### PODZIĘKOWANIA

Autorzy składają podziękowania wszystkim osobom, które przyczyniły się do powstania raportu<sup>15</sup> podsumowującego pracę badawczą, na podstawie której powstał niniejszy artykuł, w szczególności: kierownikowi projektu dyrektor Departamentu Badań Demograficznych GUS Dorocie Szałtys oraz członkom zespołu badawczego: Michałowi Adamskiemu, Mariuszowi Chmielewskiemu, Piotrowi Filipowi, Danielowi Godlewskiemu, Tomaszowi Józefowskiemu, Pawłowi Kaczorowskiemu, Zofii Kostrzewie, Jackowi Kowalewskiemu, Arlecie Olbrot-Brzezińskiej, Arturowi Owczarkowskiemu, Joannie Stańczak, Karinie Stelmach oraz Annie Wysockiej.

#### BIBLIOGRAFIA

- Bakker, B. F. M., van der Heijden, P. G. M., Gerritse, S. C. (2017). Estimation of non-registered usual residents in the Netherlands. W: D. Böhning, P. G. M. van der Heijden, J. Bunge (red.), *Capture-Recapture Methods for the Social and Medical Sciences* (s. 259–273). Boca Raton, Florida: CRC Press.
- Bouchard, M. (2007). A Capture-Recapture Model to Estimate the Size of Criminal Populations and the Risks of Detection in a Marijuana Cultivation Industry. *Journal of Quantitative Criminology*, 23(3), 221–241. DOI: 10.1007/s10940-007-9027-1.
- Bouchard, M. (2008). Towards a Realistic Method to Estimate Cannabis Production in Industrialized Countries. *Contemporary Drug Problems*, 35(2–3), 291–320. DOI: 10.1177/009145090803500206.
- Bouchard, M., Tremblay, P. (2005). Risks of Arrest across Drug Markets: A Capture-Recapture Analysis of "Hidden" Dealer and User Populations. *Journal of Drug Issues*, 35(4), 733–754. DOI: 10.1177/002204260503500404.

---

<sup>15</sup> Raport wraz z załącznikami dostępny jest na stronie <http://stat.gov.pl/statystyka-regionalna/statystyka-dla-polityki-spojnosci/statystyka-dla-polityki-spojnosci-2016-2018/badania/rynek-pracy-ubostwo-i-wykluczenie-spoeczne/>.

- Brzezińska, J. (2015). *Analiza logarytmiczno-liniowa: teoria i zastosowania z wykorzystaniem programu R*. Warszawa: Wydawnictwo C.H. Beck.
- Buckland, S. T., Garthwaite, P. H. (1991). Quantifying Precision of Mark-Recapture Estimates Using the Bootstrap and Related Methods. *Biometrics*, 47(1), 255–268. DOI: 10.2307/2532510.
- Böhning, D., van der Heijden, P. G. M., Bunge, J. (2017). *Capture-Recapture Methods for the Social and Medical Sciences*. Boca Raton, Florida: CRC Press.
- Böhning, D., van der Heijden, P. G. M., Bunge, J. (2018). Basic concepts of capture-recapture. W: D. Böhning, P. G. M. van der Heijden, J. Bunge (red.), *Capture-Recapture Methods for the Social and Medical Sciences* (s. 237–257). Boca Raton, Florida: CRC Press.
- Chao, A. (1989). Estimating Population Size for Sparse Data in Capture-Recapture Experiments. *Biometrics*, 45(2), 427–438. DOI: 10.2307/2531487.
- Chen, B., Shrivastava, A., Steorts, R. C. (2018). Unique entity estimation with application to the Syrian conflict. *The Annals of Applied Statistics*, 12(2), 1039–1067. DOI: 10.1214/18-AOAS1163.
- Coumans, A. M., Cruyff, M., van der Heijden, P. G. M., Wolf, J. R. L. M., Schmeets, H. (2017). Estimating Homelessness in the Netherlands Using a Capture-Recapture Approach. *Social Indicators Research*, 130(1), 189–212. DOI: 10.1007/s11205-015-1171-7.
- Di Cecco, D., Di Zio, M., Filippini, D., Rocchetti, I. (2018). Population Size Estimation Using Multiple Incomplete Lists with Overcoverage. *Journal of Official Statistics*, 34(2), 557–572. DOI: 10.2478/jos-2018-0026.
- Di Consiglio, L., Tuoto, T. (2015). Coverage Evaluation on Probabilistically Linked Data. *Journal of Official Statistics*, 31(3), 415–429. DOI: 10.1515/jos-2015-0025.
- Gemmell, I., Millar, T., Hay, G. (2004). Capture-recapture estimates of problem drug use and the use of simulation based confidence intervals in a stratified analysis. *Journal of Epidemiology & Community Health*, 58(9), 758–765. DOI: 10.1136/jech.2003.008755.
- Gerritse, S. C. (2016). *An application of population size estimation to official statistics: sensitivity of model assumptions and the effect of implied coverage*. Pobrane z: <https://dspace.library.uu.nl/handle/1874/337476>.
- Gerritse, S. C., van der Heijden, P. G. M., Bakker, B. F. M. (2015). Sensitivity of Population Size Estimation for Violating Parametric Assumptions in Log-linear Models. *Journal of Official Statistics*, 31(3), 357–379. DOI: 10.1515/jos-2015-0022.
- Godwin, R. T., Böhning, D. (2017). Estimation of the population size by using the one-inflated positive Poisson model. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66(2), 425–448. DOI: 10.1111/rssc.12192.
- Golata, E. (2012). Spis ludności i prawda. *Studia Demograficzne*, 161(1), 23–55. DOI: 10.2478/v10274-012-0002-y.
- Golata, E. (2018). *Koniec ery tradycyjnych spisów ludności*. Poznań: Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu.
- Goodman, L. A. (1964). Simple Methods for Analyzing Three-Factor Interaction in Contingency Tables. *Journal of the American Statistical Association*, 59(306), 319–352. DOI: 10.1080/01621459.1964.10482163.
- Goodman, L. A. (1968). The Analysis of Cross-Classified Data: Independence, Quasi-Independence, and Interactions in Contingency Tables with or without Missing Entries: R. A. Fisher memorial lecture. *Journal of the American Statistical Association*, 63(324), 1091–1131. DOI: 10.1080/01621459.1968.10480916.
- Goodman, L. A. (1969). On Partitioning  $\chi^2$  and Detecting Partial Association in Three-Way Contingency Tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 31(3), 486–498. DOI: 10.1111/j.2517-6161.1969.tb00808.x.

- Goudie, I. B. J., Goudie, M. (2007). Who captures the marks for the Petersen estimator? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(3), 825–839. DOI: 10.1111/j.1467-985X.2007.00479.x.
- Griffin, R. A. (2014). Potential uses of administrative records for triple system modeling for estimation of census coverage error in 2020. *Journal of Official Statistics*, 30(2), 177–189. DOI: 10.2478/jos-2014-0012.
- van der Heijden, P. G. M., Cruts, G., Cruyff, M. (2013). Methods for population size estimation of problem drug users using a single registration. *International Journal of Drug Policy*, 24(6), 614–618. DOI: 10.1016/j.drugpo.2013.04.002.
- van der Heijden, P. G. M., Cruyff, M., van Houwelingen, H. C. (2003). Estimating the Size of a Criminal Population from Police Records Using the Truncated Poisson Regression Model. *Statistica Neerlandica*, 57(3), 289–304. DOI: 10.1111/1467-9574.00232.
- van der Heijden, P. G. M., Whittaker, J., Cruyff, M., Bakker, B., van der Vliet, R. (2012). People born in the Middle East but residing in the Netherlands: invariant population size estimates and the role of active and passive covariates. *The Annals of Applied Statistics*, 6(3), 831–852.
- Hudson, C. G. (1998). Estimating Homeless Populations through Structural Equation Modeling. *The Journal of Sociology & Social Welfare*, 25(2), 136–154.
- Jędrzejczak, A., Kubacki, J. (2014). Problemy jakości danych statystycznych w przypadku badania cech rzadkich. *Wiadomości Statystyczne*, (6), 11–26.
- Rossi, C., Mascioli, F. (2008). Capture-recapture methods to estimate prevalence indicators for evaluating drug policies. *Bulletin on Narcotics*, 15(1–2), 5–25.
- Schepers, W., Nicaise, I. (2017). Estimating the homeless population. Sampling strategies. (HIVA Working Paper Series No. 20). Pobrane z: [https://www.belspo.be/belspo/brain-be/projects/FinalReports/MEHOBEL\\_Final%20report\\_Annex%20I.pdf](https://www.belspo.be/belspo/brain-be/projects/FinalReports/MEHOBEL_Final%20report_Annex%20I.pdf).
- The R Development Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Tourangeau, R., Edwards, B., Johnson, T. P., Wolter, K. M., Bates, N. (2014). *Hard-to-Survey Populations*. Cambridge: Cambridge University Press.
- Wolter, K. M. (1986). Some Coverage Error Models for Census Data. *Journal of the American Statistical Association*, 81(394), 337–346. DOI: 10.1080/01621459.1986.10478277.
- Zhang, L.-C. (2008). *Developing methods for determining the number of unauthorized foreigners in Norway*. Oslo: Statistisk Sentralbyrå. Pobrane z: [https://www.ssb.no/a/english/publikasjoner/pdf/doc\\_200811\\_en/doc\\_200811\\_en.pdf](https://www.ssb.no/a/english/publikasjoner/pdf/doc_200811_en/doc_200811_en.pdf).
- Zhang, L.-C. (2015). On Modelling Register Coverage Errors. *Journal of Official Statistics*, 31(3), 381–396. DOI: 10.1515/jos-2015-0023.
- Zhang, L.-C., Dunne, J. (2018). Trimmed dual system estimation. W: D. Böhning, P. G. M. van der Heijden, J. Bunge (red.), *Capture-recapture methods for the social and medical sciences* (s. 237–257). Boca Raton, Florida: CRC Press.
- Zwane, E. N., van der Heijden, P. G. M. (2003). Implementing the parametric bootstrap in capture-recapture models with continuous covariates. *Statistics & Probability Letters*, 65(2), 121–125. DOI: 10.1016/j.spl.2003.07.010.
- Zwane, E., van der Heijden, P. G. M. (2005). Population estimation using the multiple system estimator in the presence of continuous covariates. *Statistical Modelling*, 5(1), 39–52. DOI: 10.1191/1471082X05st086oa.

## Impact of joint-stock companies' financial condition on real activities manipulation to manage earnings

Sergiusz Herman<sup>a</sup> 

**Abstract.** The aim of the paper is both to determine whether joint-stock companies in poor financial condition undertake real activities manipulation to manage earnings, and to investigate the potential relationship between the scope of real activities manipulation and business fields in which these companies are operating. The database of the Notoria Serwis company and panel data models were used for the purpose of the study. In order to measure the scale of real earnings manipulation, the author adopted Roychowdhury's methodology, focusing on the manipulation of sales, production costs and discretionary expenses.

The study examined a sample of non-financial joint-stock companies listed on the Warsaw Stock Exchange in the period 1998–2016, and was based on 1493–1669 observations. The results demonstrated that companies in poor financial condition undertake actions to manage earnings. Managers manipulate sales volumes and reduce levels of discretionary expenses. The intensity of those actions depends on the type of business activity of a given company.

**Keywords:** joint-stock companies, earnings management, financial condition, real activity manipulation

## Wpływ kondycji finansowej spółek akcyjnych na podejmowanie działań o charakterze realnym w celu zarządzania zyskami

**Streszczenie.** Celem artykułu jest, po pierwsze, ustalenie, czy spółki akcyjne będące w gorszej kondycji finansowej podejmują działania o charakterze realnym w celu zarządzania zyskami, a po drugie, zweryfikowanie, czy skala podejmowanych działań zależy od rodzaju działalności gospodarczej przedsiębiorstwa. W badaniu wykorzystano bazę danych firmy Notoria Serwis oraz modele panelowe. Dla określenia skali działalności związanej z manipulowaniem zyskami posłużono się metodyką zaproponowaną przez Roychowdhury, koncentrując się na manipulacjach związanych ze sprzedażą, produkcją oraz kosztami pośrednimi.

Badanie przeprowadzono na podstawie próby niefinansowych spółek akcyjnych funkcjonujących na Giełdzie Papierów Wartościowych w Warszawie w latach 1998–2016; obejmowało łącznie od 1493 do 1669 obserwacji. Wyniki pozwalają stwierdzić, że spółki będące w gorszej kondycji finansowej podejmują działania o charakterze realnym związane z zarządzaniem zyskami. Zarządzający manipulują wielkością sprzedaży oraz redukują koszty pośrednie produkcji. Natężenie tych działań jest związane z rodzajem działalności gospodarczej prowadzonej przez dany podmiot.

**Słowa kluczowe:** spółki akcyjne, zarządzanie zyskami, kondycja finansowa, działania o charakterze realnym

**JEL:** M41, G30, C33

---

<sup>a</sup> Poznań University of Economics and Business, Department of Econometrics.

Earnings management is a well-known subject, described in literature. Its popularity grew fast after the financial crisis on the US market in the early 2000s.

According to literature, earnings management occurs when the management of a company decides to use accounting techniques which make it possible to change the contents of the company's financial statements:

- to give stakeholders a false impression of the company's financial condition;
- to fulfil the terms of other contracts that depend on reported statements (Healy & Wahlen, 1999).

According to Healy and Wahlen (1999), it is the management's decision, based on their judgement, to adopt such techniques or not. It is worth noting that such practices are allowed by the law. Various actions might be taken to manage earnings. They can be divided into two basic categories (Roychowdhury, 2006; Zang, 2011). The first is accrual-based earnings management, which involves transaction recognition in financial statements. This kind of activity does not directly affect the level of cash flow. Examples include changing the depreciation method for fixed assets or delaying impairment losses of fixed assets. Accrual-based earnings management results in a changed financial statement without making any new transactions.

The second category is real activities manipulation which can be defined as actions of the management distinctly different from standard business practices, whose purpose is to obtain a certain financial result. Real activities manipulation affects the level of cash flow. Examples include decreasing expenditures on research and development or offering high discounts to customers.

Real activities manipulation is the subject of the study presented in this paper. The study focused on three types of activities from that category (Roychowdhury, 2006):

- sales manipulation – activity aimed at temporarily increasing sales revenues in a given year, e.g. by offering discounts or more favourable financing terms to customers. Such actions result in achieving a higher profit at the end of the accounting period, but on the other hand, they negatively affect the level of cash flow, which is consequently lower than expected for a certain level of sales volume;
- reduction of discretionary expenses – limiting expenditure on research and development or cutting advertising or general management costs. As a result, the company is able to present a higher profit in its financial statement. The level of indirect costs is consequently lower than expected for a certain level of sales volume;
- overproduction – production increased to the level exceeding demand. For a higher production level, fixed costs are divided between more items, which translates into a lower unit cost. As a result, the company has higher operating margins, but also higher costs connected to the production and storage of products. Consequently, the company's production costs amount to the level higher than expected.

The management can undertake each of the above-mentioned activities to achieve a certain financial result. However, in the long run, such practices may reduce the value of a company. Overstating financial results in a given accounting period is bound to have a negative effect on the level of cash flow in subsequent periods.

Roychowdhury (2006) wrote one of the most important publications on real activities manipulation. His methodology was used in a number of publications, most of which concerned the US capital market. Real earnings manipulation was studied there in relation to: changes in the law (Cohen, Dey & Lys, 2008), legal forms of companies (Haga, Höglund & Sundvik, 2018), the firm value (Mellado-Cid, Jory & Ngo, 2008), the auditing company's reputation (Magnis & Iatridis, 2017) and the professional experience of the management (Jiang, Zhu & Huang, 2013).

Since there have only been few studies on this subject concerning the Polish capital market, the author decided to conduct an empirical study in this area based on the performance of joint-stock companies listed on the Warsaw Stock Exchange (WSE).

The main goal of the paper is to determine whether joint-stock companies in poor financial condition undertake real activities manipulation to manage earnings. The findings of the study are also expected to answer the question whether the scope of real activities manipulation depends on a company's business profile.

The article consists of four sections. The first section presents a set of empirical studies pertaining to earnings management on the Polish capital market. The second section explores the relationship between the financial condition of companies and earnings management. The research methodology and the research sample are presented in the third section of the article, and the results obtained in the study and their interpretation in the fourth section.

## EARNINGS MANAGEMENT ON THE POLISH CAPITAL MARKET

Polish literature on earnings management includes a series of papers which can be divided into three groups.

The first group comprises empirical studies which analyse the reported financial results and their distribution. The authors Jackowicz and Kuryłek used a research sample made up of commercial banks in Poland in the years 1994–2003 in their study (Jackowicz and Kuryłek, 2005). They demonstrated a significant discontinuity in the threshold set by the zero financial result. The correlation was stronger in the case of banks with a comparatively low capital endowment, and during the period of low economic growth. Interestingly, such activities were not highly popular among banks listed on the stock exchange.

Later, this study was extended to commercial banks in Central and Eastern Europe (Jackowicz & Kozłowski, 2010). The research sample consisted of

commercial banks in 11 European countries in the years 1991–2008. The results showed that while banks avoided reporting losses, they did not seem to counteract the decrease of profitability. The results did not change when the sample was divided according to the geographic location, the quality of capital equipment, the company's presence on the stock exchange, or the macroeconomic situation of the country of origin.

Another set of studies analysing financial result distribution was conducted by Wójtowicz. In the first publication (Wójtowicz, 2007a), the author tried to find out whether earnings management could be effectively checked by audits. The research sample consisted of companies listed on the WSE in the years 1998–2004. The analysis of the distribution of net financial result and cash flow in the operational activity showed that earnings management undertaken in order to avoid losses was not significantly limited by ongoing audits. There were ways to manage earnings so that a statutory auditor was not able to stop this procedure.

In the following study (Wójtowicz, 2007b), the author attempted to determine whether the scope of earnings management depended on the legal form of a company. The research, based on 1078 observations from the period 2003–2004, showed that earnings management to avoid losses occurred in nearly 28% of limited liability companies. In the case of general limited and other partnerships as well as sole proprietorships, no indication of earnings management was detected.

Another study by Wójtowicz (2012) was aimed at verifying the hypothesis that economic climate determines the scope of financial result management to recover losses. The study was based on 2229 observations of public companies in the period 1997–2010, and analysed breakdowns of net financial results and cash flow from operating activities. The results provided insufficient evidence to reject the hypothesis. However, it was demonstrated that in good economic times, earnings management does not occur. The situation changes significantly when the economic situation is deteriorating. When the actual financial result of a company is unsatisfactory for the management, they will be trying to shape it according to their expectations.

Further research by the same author (Wójtowicz, 2015) attempted to detect any signs of earnings management to achieve zero or small positive earnings surprises. The research sample consisted of 609 observations of mWIG400 index companies in the years 2012–2014. When the author analysed the distribution of scaled annual earnings surprises (difference between actual and forecasted profits), he observed a high frequency of small positive surprises, i.e. actual profits being slightly higher than forecasted profits. This indicates that earnings management is likely to have taken place there.

Brzeszczyński, Gajdka and Schabek (2011) author the latest study on earnings management where breakdowns of financial results were analysed. As the other above-mentioned studies, their research focused on earnings distribution,



and more specifically, on earnings per share and their change over time. The sample consisted of 359 companies listed on the WSE in the period 1997–2010. This empirical study indicated that Polish companies manage earnings at the level of around zero threshold. However, it concerns the “zero growth” threshold to a lesser extent. The results of the study showed that companies which adopted earnings manipulation achieved worse results in the consecutive periods than the companies which did not undertake such activity.

The second group of studies concerning the Polish capital market focuses on the estimation of accruals in order to identify accrual-based earnings management activities. A comprehensive study in this area was conducted by Piosik (2013). The research sample was comprised of companies listed on the WSE in the period 2006–2009. The author carried out a comparative analysis of models describing the change in the value of current nonmonetary assets and studied the influence of changing accounting standards on earnings management in Poland. The results indicated that, among others, there were differences in the scale of earnings management between the companies applying the Accounting Act and those applying the International Financial Reporting Standards.

Further studies connected to the estimation of accruals were conducted by Lizińska (2015) and Lizińska and Czapiewski (2016). Both papers attempted to determine whether companies launching initial public offering (IPO) were at the same time managing earnings. The first study examined companies launching IPOs on developed and developing markets in the years 2005–2011. The second focused on companies from nonfinancial sectors launching IPOs on the WSE in years 1998–2013. Both studies demonstrated that earnings management took place during issuance periods.

The third group of studies pertaining to the Polish capital market comprises works which attempt to identify real activities manipulation. Three empirical studies should be mentioned here. The first of them was carried out by Kałdoński and Jewartowski (2017). The research examined the potential connection between equity overvaluation and various types of earnings management (both accrual- and real-based). The research sample consisted of 356 nonfinancial joint-stock companies listed on the WSE in the period 2005–2015. The results indicated that the overvaluation of equity led to the intensification of accrual-based activities, but did not induce real-based activities.

In another research (Kałdoński & Jewartowski, 2018) the same authors tried to find out whether there was a relationship between institutional ownership stability and the level of both real and accrual-based earnings management. The research, performed on a sample of 217 companies (1204 observations) listed on the WSE in the period 2007–2016, showed that such a relationship existed, but only with regard to real activities manipulation.

The most recent research investigating real activities manipulation was carried out by Kałdoński, Jewartowski and Mizerka (in press). The aim of their study

was to determine whether institutional ownership stability influences the use of real-based activities. The research sample consisted of 1204 observations of non-financial joint-stock companies listed on the WSE in the period 2007–2016. The research demonstrated that the presence of institutional investors considerably limited real earnings management, especially in the case of firms whose managers were pressurised by capital market forces to “meet or beat” earnings targets. The opposite situation took place in companies whose managers were protected from the capital market pressure.

The literature review of empirical research pertaining to the Polish capital market yielded some interesting results. The vast majority of the presented studies employed the method of analysing the distribution of companies' financial results. Only a few of the above-mentioned works attempted to identify real and accrual-based earnings management on the Polish capital market.

#### EARNINGS MANAGEMENT AND COMPANIES' FINANCIAL CONDITION

Earlier literature dealing with the relationship between the financial condition of companies and real earnings management is rather limited. The first paper on this subject, authored by Charitou, Lambertides and Trigeorgis (2007a), presented a study based on a research sample of 455 US companies that went bankrupt between the years 1986–2001, and the same number of companies in good financial condition. The results showed that earnings management occurred a year before filing for bankruptcy, and that companies manipulated mainly current accruals.

The same authors extended their analysis to another study (Charitou, Lambertides & Trigeorgis, 2007b). The research sample was made up of 859 companies that went bankrupt in the years 1986–2004. The aim of the study was to identify the motivation behind the managements' decisions to use earnings management. The results showed that what had a big impact on such decisions were the opinions of statutory auditors, but they also depended on the companies' ownership structure.

Further study on the relationship between companies' financial condition and earnings management was conducted by García Lara, Osma and Neophytou (2009). The study examined 268 British companies that went bankrupt in the years 1998–2004. Its results demonstrated that companies which filed for bankruptcy also undertook earnings management activities. Such activities were observed even as early as four years before the bankruptcy. The managements tried to hide the deteriorating financial condition of their companies, and they engaged in both accrual-based earnings management and real activities manipulation.

A similar study was carried out for the New Zealand market (Habib, Uddin Bhuiyan & Islam, 2013). The research sample consisted of 767 companies ob-

served for a year, in the period 2000–2011. The results confirmed the findings of other studies on this subject, namely that managements of companies at risk of bankruptcy manage earnings. The study also indicated that the global economic crisis had a considerable impact on managements' decisions.

Campa and Camacho-Miñano's analysis (2015) also explored the relationship between companies' financial condition and the use of real activities manipulation. The authors attempted to find out which tools for earnings management were used by the managements of companies at risk of bankruptcy. The research sample consisted of 362 companies that went bankrupt in Spain. According to the results, these companies indeed engaged in earnings management. If a company was in poor financial condition, the management was more likely to use real activities manipulation. It was not checked, however, whether the intensity of these activities depended on the business profile of a given company.

The literature review of the relationship between the financial condition of a company and earnings management showed that all the studies pertaining to this issue were comparing a group of companies that went bankrupt with companies in good financial condition. The analyses, then, concerned only the two extreme situations a company can experience. Therefore, the author decided to conduct an empirical study in which an intermediate situation is being considered, namely, he posed a question whether there was any relationship between companies' deteriorating financial condition and the decisions to employ earnings management.

## METHODOLOGY

According to literature, earnings management results in cash flow and discretionary expenses lower than expected for a given sales volume, as well as surprisingly high production costs. In order to estimate the values that are regarded as typical or expected, Roychowdhury's (2006) methodology was used. The following model was adopted to estimate the normal level of cash flow:

$$\frac{CFO_{it}}{ASSETS_{it-1}} = \alpha + \beta_{1t} \left( \frac{1}{ASSETS_{it-1}} \right) + \beta_{2t} \left( \frac{SALES_{it}}{ASSETS_{it-1}} \right) + \beta_{3t} \left( \frac{\Delta SALES_{it}}{ASSETS_{it-1}} \right) + \varepsilon_{it} \quad (1)$$

where:

$CFO_{it}$  – cash flow from operations in period  $t$ ,

$ASSETS_{it-1}$  – total assets at the end of period  $t - 1$ ,

$SALES_{it}$  – sales during period  $t$ ,

$\Delta SALES_{it} = SALES_{it} - SALES_{it-1}$ .

There is a linear relationship between the level of cash flow from operating activities, estimated in the above-mentioned way, and the revenues and their change. The level of abnormal cash flow (*Abnormal CFO*) for each company and studied year is the difference between the actual level of cash flow and the one estimated following the above model. An indicator of earnings manipulation in companies in poor financial condition is a lower level of abnormal cash flow than in companies in good financial condition.

As mentioned above, overproduction in a company results in the level of production costs higher than expected. According to Roychowdhury (2006), production costs are the sum of the cost of goods sold and the change in the inventory. The normal level of production costs was estimated according to the model below:

$$\begin{aligned} \frac{PROD_{it}}{ASSETS_{it-1}} = & \alpha + \beta_{1t} \left( \frac{1}{ASSETS_{it-1}} \right) + \beta_{2t} \left( \frac{SALES_{it}}{ASSETS_{it-1}} \right) + \\ & + \beta_{3t} \left( \frac{\Delta SALES_{it}}{ASSETS_{it-1}} \right) + \beta_{4t} \left( \frac{\Delta SALES_{it-1}}{ASSETS_{it-1}} \right) + \varepsilon_{it} \end{aligned} \quad (2)$$

where:

- $PROD_{it}$  – production costs in period  $t$ ,  
 $ASSETS_{it-1}$  – total assets at the end of period  $t - 1$ ,  
 $SALES_{it}$  – sales during period  $t$ ,  
 $\Delta SALES_{it} = SALES_{it} - SALES_{it-1}$ ,  
 $\Delta SALES_{it-1} = SALES_{it-1} - SALES_{it-2}$ .

Therefore, the level of production costs depends on revenues and changes in them over time. The level of abnormal production costs (*Abnormal PROD*) for each company and studied year is the difference between the actual production costs and their estimation obtained following the above model. The indicator of earnings manipulation in companies in poor financial condition is a higher level of abnormal production costs than in companies in good financial condition.

As stated in the introduction, the reduction of discretionary expenses in a company results in their abnormal, lower than expected, level. According to the methodology used in this study, discretionary expenses depend on the level of revenues, which can be described by the model below:

$$\frac{DISEXP_{it}}{ASSETS_{it-1}} = \alpha + \beta_{1t} \left( \frac{1}{ASSETS_{it-1}} \right) + \beta_{2t} \left( \frac{SALES_{it-1}}{ASSETS_{it-1}} \right) + \varepsilon_{it} \quad (3)$$

where:

- $DISEXP_{it}$  – discretionary expenses in period  $t$ ,  
 $ASSETS_{it-1}$  – total assets at the end of period  $t - 1$ ,  
 $SALES_{it-1}$  – sales during period  $t - 1$ .

The level of abnormal discretionary expenses (*Abnormal DISEXP*) for each company and studied year was estimated as the difference between the actual level of discretionary expenses and their calculation based on the model. As in the case of cash flow, the indicator of earnings manipulation in companies in poor financial condition is a lower value of abnormal discretionary expenses than in companies in good financial condition. The presented models (1)–(3) were estimated separately with the use of the ordinary least squares method for each combination of business activity and studied year.

The main aim of this paper is to determine whether joint-stock companies in poor financial condition undertake real activities manipulation to manage earnings. To define the financial condition of a company in every studied year, a composite measure was used (Panek, 2009). The following financial indicators<sup>1</sup> (with preferred value range for nominants<sup>2</sup>) were used to create the measure:

- return on assets;
- return on equity;
- quick ratio (0.8–1.2);
- cash flow liquidity (0.1–0.2);
- receivables turnover ratio (7–17);
- inventory turnover (7–17);
- debt ratio (0.57–0.60).

After converting all the above indicators to stimulants, their values were normalized using the zero unitarisation method<sup>3</sup>, thanks to which they assumed non-negative values. Next, a composite measure was calculated as the arithmetic mean of the standardised values. The values of the measure reflected the financial condition of studied companies (measure referred to as *CONDITION*).

An adequate research sample was essential to estimate models (1)–(3). The database of the Notoria Serwis company was used for this purpose. The research examined yearly financial statements of companies operating in the years 1998–2016 in six business fields (the fields with the strongest representation on the WSE):

- food;
- construction;
- IT;
- real estate;
- textiles and cosmetics;
- mechanical and electrical engineering.

---

<sup>1</sup> Coefficients of variation for all variables were higher than 50%.

<sup>2</sup> Based on financial literature.

<sup>3</sup> It was assumed that profitability ratios are stimulants while the others are nominants.

The study examined companies with financial data available for at least two consecutive years, and those years for which the data of at least 15 companies was available were analysed. Taking all the above into account, the author managed to gather the following numbers of company-year observations: 1669 for the first model (1), 1383 for the second model (2), and 1493 for the third model (3). To eliminate all outliers, all financial data were winsorized at the 5th and 95th percentile.

As mentioned above, companies in poor financial condition using real activities manipulation were expected to demonstrate a lower level of abnormal cash flow and discretionary expenses than companies in good financial condition, and the opposite results were anticipated in the case of abnormal production costs. To prove the above assertion, the following panel data models were estimated:

$$Y_{it} = \alpha + \beta_1 \text{CONDITION}_{it} + \beta_2 \text{SIZE}_{it} + \beta_3 \text{LEV}_{it} + \beta_4 \text{GROWTH}_{it} + \beta_5 \text{EISSUE}_{it} + \beta_6 \text{DISSUE}_{it} + \varepsilon_{it} \quad (4)$$

where:

$Y$  – Abnormal CFO, Abnormal PROD or Abnormal DISEXP,

$\text{CONDITION}_{it}$  – composite measure (financial condition) at the end of period  $t$ ,

$\text{SIZE}_{it}$  – natural logarithm of total assets at the end of period  $t$ ,

$\text{LEV}_{it}$  – leverage measured as total liabilities over total equity at the end of period  $t$ ,

$\text{GROWTH}_{it}$  – annual change in net sales,

$\text{EISSUE}_{it}$  – annual change in shareholder equity,

$\text{DISSUE}_{it}$  – annual change in liabilities.

Abnormal cash flow, discretionary expenses and production costs estimated with the use of (1)–(3) models served as dependent variables in the estimated panel models. Apart from the independent variable showing the financial condition of companies ( $\text{CONDITION}_{it}$ ), control variables (following earlier literature, e.g. Campa, 2015; Haga et al., 2018) were introduced to the models, because previous studies showed that their level may significantly impact the scope of earnings management activities in a company. Therefore, the following additional variables were introduced to the model: the size of a company (Park & Shin, 2004), leverage (Dechow, Sloan & Sweeney, 1995), growth of sales volume (Carey & Simnett, 2006) and the way of financing (Shan, Taylor & Walter, 2010).

## RESULTS

Firstly, the strength of the correlation between all the variables of the model (4) was analysed. Table 1 presents the Pearson correlation coefficients for all analysed variables.

TABLE 1. PEARSON CORRELATION MATRIX

Variable	<i>Abnormal CFO</i>	<i>Abnormal PROD</i>	<i>Abnormal DISEXP</i>	<i>SIZE</i>	<i>LEV</i>	<i>GROWTH</i>	<i>EISSUE</i>	<i>DISSUE</i>	<i>CONDITION</i>
<i>Abnormal CFO</i>	1.00	-0.09	0.03	0.00	-0.03	0.00	0.01	-0.47	0.13
<i>Abnormal PROD</i> .....	-0.09	1.00	-0.02	-0.04	0.05	0.15	0.09	0.10	0.03
<i>Abnormal DISEXP</i> .....	0.03	-0.02	1.00	0.01	0.01	0.19	0.08	0.01	0.05
<i>SIZE</i> .....	0.00	-0.04	0.01	1.00	0.22	0.04	0.00	0.00	-0.02
<i>LEV</i> .....	-0.03	0.05	0.01	0.22	1.00	0.06	-0.04	0.00	-0.35
<i>GROWTH</i> .....	0.00	0.15	0.19	0.04	0.06	1.00	0.17	0.01	0.18
<i>EISSUE</i> .....	0.01	0.09	0.08	0.00	-0.04	0.17	1.00	0.03	0.12
<i>DISSUE</i> .....	-0.47	0.10	0.01	0.00	0.00	0.01	0.03	1.00	0.01
<i>CONDITION</i> ....	0.13	0.03	0.05	-0.02	-0.35	0.18	0.12	0.01	1.00

Note. Statistically significant coefficients are in italics.

Source: own calculation based on the Notoria Serwis database.

According to the results displayed in Table 1, the levels of abnormal cash flow (*Abnormal CFO*) and abnormal discretionary expenses (*Abnormal DISEXP*) have a positive correlation coefficient with the financial condition of companies. These correlations are statistically significant (at the significance level of 0.05). It means that the lower the value of the composite measure, the worse the financial condition of the studied companies, and the lower the level of abnormal cash flow and discretionary expenses. Therefore, as mentioned in the second section of the paper, in such cases the managements are likely to use real activities manipulation to manage earnings. In the case of the *Abnormal PROD* variable, a statistically significant correlation was not observed. Furthermore, the data presented in Table 1 indicates that the majority of control variables are strongly correlated with variables describing abnormal cash flow, discretionary expenses and production costs. For that reason, a correlation between the financial condition and the scale of earnings management of a company should be studied prior to the use of a multidimensional approach. Taking the results of Breusch-Pagan and Hausman tests into account, relevant panel data models (4) were estimated. The first model presents random effects, and the other two fixed effects. The results are presented in Tables 2–4.

In the case of the panel data model where the *Abnormal CFO* (Table 2) was the dependent variable, only three independent variables were statistically significant (at the significance level of 0.05). One of them was the variable reflecting the financial condition of the studied joint-stock companies, and, what is important, its parameter had a positive value. It means that companies in poor financial condition present a lower level of abnormal cash flow from operating activity (for a given sales volume) than companies in good financial condition. This indicates that the managements of those companies might manipulate sales volumes.

**TABLE 2. PANEL DATA MODEL FOR A DEPENDENT VARIABLE ABNORMAL CFO**

Specification	Coeff	Std. error	t-statistic	p-value
<i>Const</i> .....	-0.1300	0.0509	-2.557	0.011
<i>CONDITION</i> .....	0.3915	0.0999	3.915	0.000
<i>SIZE</i> .....	-0.0001	0.0031	-0.034	0.973
<i>LEV</i> .....	0.0046	0.0032	1.404	0.160
<i>GROWTH</i> .....	0.0001	0.0086	0.008	0.994
<i>EISSUE</i> .....	0.0010	0.0015	0.651	0.515
<i>DISSUE</i> .....	-0.0036	0.0002	-23.180	0.000
Mean dependent var. ....			-0.003	
Sum squared resid. ....			19.422	
Log-likelihood .....			1348.284	
Schwarz criterion .....			-2644.627	
S.D. dependent var. ....			0.124	
S.E. of regression .....			0.108	
Akaike criterion .....			-2682.567	
Hannan-Quinn .....			-2668.509	
Breusch-Pagan test .....		302.021		0.000
Hausman test .....		11.647		0.070

S o u r c e: as in Table 1.

In the model where the *Abnormal PROD* was a dependent variable (Table 3), four variables were statistically significant (at the significance level of 0.05). These were the variables reflecting the financial condition of a company, the leverage and the growth of sales volume. Moreover, the sign of the parameter for the financial condition variable suggests that the level of production costs is lower for companies in poor financial condition. It is not an indicator of earnings management in a company.

**TABLE 3. PANEL DATA MODEL FOR A DEPENDENT VARIABLE ABNORMAL PROD**

Specification	Coeff	Std. error	t-statistic	p-value
<i>Const</i> .....	-1.1531	0.3878	-2.974	0.003
<i>CONDITION</i> .....	1.6233	0.5135	3.161	0.002
<i>SIZE</i> .....	0.0304	0.0271	1.257	0.209
<i>LEV</i> .....	0.0533	0.0186	2.864	0.004
<i>GROWTH</i> .....	0.1835	0.0429	4.272	0.000
<i>EISSUE</i> .....	0.0053	0.0077	0.680	0.497
<i>DISSUE</i> .....	-0.0321	0.0183	-1.752	0.080
Mean dependent var. ....			0.033	
Sum squared resid. ....			200.608	
LSDV R-squared .....			0.294	
LSDV F(174, 1494) .....			3.133	
Log-likelihood .....			-627.341	
Schwarz criterion .....			2433.499	
S.D. dependent var. ....			0.453	
S.E. of regression .....			0.406	
Within R-squared .....			0.034	
p-value (F) .....			0.000	
Akaike criterion .....			1580.681	
Hannan-Quinn .....			1899.678	

S o u r c e: as in Table 1.



The analysis of the results of the panel data model in which the *Abnormal DISEXP* was the dependent variable (Table 4) showed that six variables were statistically significant (at the significance level of 0.1). One of them was the variable reflecting the financial condition of the studied joint-stock companies. A positive value of the parameter for this variable means that the level of discretionary expenses is lower for companies in poor financial condition. It indicates that their managements might have reduced the level of these expenses to manage earnings.

**TABLE 4. PANEL DATA MODEL FOR A DEPENDENT VARIABLE *ABNORMAL DISEXP***

Specification	Coeff	Std. error	t-statistic	p-value
<i>Const</i> .....	-0.4927	0.1068	-4.615	0.000
<i>CONDITION</i> .....	0.2682	0.1465	1.831	0.067
<i>SIZE</i> .....	0.0215	0.0071	3.020	0.003
<i>LEV</i> .....	0.0084	0.0051	1.653	0.099
<i>GROWTH</i> .....	0.1153	0.0118	9.776	0.000
<i>EISSUE</i> .....	0.0036	0.0021	1.764	0.078
<i>DISSUE</i> .....	0.0001	0.0002	0.579	0.562
Mean dependent var. ....		0.009		
Sum squared resid. ....		21.008		
LSDV R-squared .....		0.525		
LSDV F(174, 1494) .....		9.145		
Log-likelihood .....		1065.552		
Schwarz criterion .....		-947.011		
S.D. dependent var. ....		0.172		
S.E. of regression .....		0.126		
Within R- squared .....		0.093		
p-value (F) .....		0.000		
Akaike criterion .....		-1807.103		
Hannan-Quinn .....		-1486.624		

Source: as in Table 1.

Another aim of this paper is to investigate the potential relationship between the scope of real activities manipulation and business fields in which the studied joint-stock companies were operating. Panel data models (Table 4) for each business field were estimated for that purpose. As a result, 24 panel data models were created. As in the case of the previous models, the results of the Breusch-Pagan and Hausman tests were taken into account here. Due to the large number of models, Table 5 presents only the information on parameter values for each business activity and the three dependent variables.

The analysis of the results of the panel data models connected with abnormal cash flow (Table 5) indicates that the parameter for the *CONDITION* variable had a positive value and was statistically significant for three business fields: food, construction, and textiles and cosmetics. For the IT, mechanical and electrical engineering and real estate sectors, the parameter did not statistically sig-

nificantly differ from zero. There were no indicators of sales volume manipulation in any of these three business fields.

In the case of the panel data models where the level of *Abnormal PROD* was a dependent variable (Table 5), the financial *CONDITION* variable was statistically significant for the IT, textiles and cosmetics and the mechanical and electrical engineering sectors, but the parameter had a negative value only for the mechanical and electrical engineering sector. This might indicate that earnings management activities were undertaken by managements in the latter sector. Interestingly, the analysis for all the examined companies studied as a group did not yield similar results.

**TABLE 5. PANEL DATA MODELS FOR EACH BUSINESS ACTIVITY AND A DEPENDENT VARIABLES *ABNORMAL CFO*, *ABNORMAL PROD* AND *ABNORMAL DISEXP***

Specification	Food	Construction	IT	Textiles and cosmetics	Mechanical and electrical engineering	Real estate
<b><i>ABNORMAL CFO</i></b>						
<i>Const</i>	-0.306**	-0.058	-0.276*	0.194	0.007	-0.204
<i>CONDITION</i>	0.534***	0.333*	0.527	0.496*	0.118	0.099
<i>SIZE</i>	0.013*	-0.005	0.007	-0.028*	0.002	0.018
<i>LEV</i>	0.009	-0.002	0.019	0.005	0.005	-0.020**
<i>GROWTH</i>	-0.002	-0.008	-0.041	-0.006	-0.013	-0.007
<i>EISSUE</i>	-0.009	-0.006	0.034***	0.003	-0.008	-0.028***
<i>DISSUE</i>	-0.029***	0.014	0.009	-0.007	0.003	-0.004***
Model		random effects			fixed effects	
<b><i>ABNORMAL PROD</i></b>						
<i>Const</i>	-0.566*	0.571	-10.264***	-1.087***	0.664**	1.866***
<i>CONDITION</i>	-0.082	-0.023	8.420***	1.314**	-1.452***	-0.142
<i>SIZE</i>	0.042*	-0.068***	0.613***	0.038	-0.022	-0.141***
<i>LEV</i>	0.022*	0.042**	0.096	0.003	-0.008	-0.015
<i>GROWTH</i>	-0.006	0.089*	0.731**	0.109**	0.009	0.037
<i>EISSUE</i>	0.020*	0.093***	-0.170	0.001	0.039**	0.015**
<i>DISSUE</i>	0.025	0.054**	-0.161**	0.089***	0.028*	0.039*
Model		fixed effects			fixed effects	
<b><i>ABNORMAL DISEXP</i></b>						
<i>Const</i>	-0.172	-0.048	-0.767***	-1.484***	-0.141*	0.287***
<i>CONDITION</i>	-0.307	-0.135	0.027	-0.168	0.028*	-0.071
<i>SIZE</i>	0.004	0.003	0.051**	0.103***	0.006	-0.020***
<i>LEV</i>	0.025***	-0.012***	-0.009	0.019	-0.007	-0.004
<i>GROWTH</i>	0.027	0.037***	-0.043	0.230***	0.036***	0.022***
<i>EISSUE</i>	0.066***	-0.001	0.055***	-0.021***	0.006	0.000
<i>DISSUE</i>	0.078***	0.025***	0.124***	0.062***	0.022***	0.000**
Model		random effects	fixed effects	random effects		fixed effects

Note. Variable is significant at the level of: \*  $\alpha = 0.1$ , \*\*  $\alpha = 0.05$ , \*\*\*  $\alpha = 0.01$ .

Source: as in Table 1.

With regard to the panel data models in which the *Abnormal DISEXP* variable was a dependent variable (Table 7), the parameter had a positive value and was statistically significant only for the mechanical and electrical engineering sector. This indicates that earnings management could have occurred only in that sector. The results suggest that discretionary expenses might have been intentionally reduced there to achieve certain financial results.

## CONCLUSIONS

The study examined non-financial joint-stock companies listed on the Warsaw Stock Exchange in the years 1998–2016. The results show that managers of joint-stock companies in poor financial condition operating in Poland use real activities manipulation to manage earnings. They manipulate the sales volume and reduce the level of discretionary expenses.

The scope of actions undertaken to manage earnings depends on the type of business activity. The manipulation of sales volume was observed in the manufacturing sector, while the reduction of discretionary expenses and overproduction in the mechanical and electrical engineering sector.

This empirical analysis should be continued in two directions. Firstly, it should attempt to identify and characterise companies that are likely to be involved in earnings management. Secondly, it should try to find out whether earnings manipulation influences the future financial condition and share quotes of a given company.

## REFERENCES

- Brzeszczyński, J., Gajdka, J., Schabek, T. (2011). Earnings management in Polish companies. *Comparative Economic Research*, 14(3), 137–150.
- Campa, D. (2015). The impact of SME's pre-bankruptcy financial distress on earnings management tools. *International Review of Financial Analysis*, 42, 222–234.
- Carey, P., Simnett, R. (2006). Audit partner tenure and audit quality. *The Accounting Review*, 81(3), 653–676.
- Charitou, A., Lambertides, N., Trigeorgis, L. (2007a). Managerial discretion in distressed firms. *The British Accounting Review*, 39(4), 323–346.
- Charitou, A., Lambertides, N., Trigeorgis, L. (2007b). Earnings behaviour of financially distressed firms: the role of institutional ownership. *Abacus*, 43(3), 271–296.
- Cohen, D. A., Dey, A., Lys, T. Z. (2008). Real and accrual-based earnings management in the pre- and post-Sarbanes-Oxley periods. *The Accounting Review*, 83(3), 757–787.
- Dechow, P. M., Sloan, R. G., Sweeney, A. P. (1995). Detecting earnings management. *The Accounting Review*, 70(2), 193–225.
- García Lara, J. M., Osmá, B. G., Neophytou, E. (2009). Earnings quality in ex-post failed firms. *Accounting and Business Research*, 39(2), 119–138.

- Habib, A., Uddin Bhuiyan, B., Islam, A. (2013). Financial distress, earnings management and market pricing of accruals during the global financial crisis. *Managerial Finance*, 39(2), 155–180.
- Haga, J., Höglund, H., Sundvik, D. (2018). Stock market listing status and real earnings management. *Journal of Accounting and Public Policy*, 37(5), 420–435.
- Healy, P. M., Wahlen, J. M. (1999). A review of the earnings management literature and its implications for standard setting. *Accounting Horizons*, 13(4), 365–383.
- Jackowicz, K., Kozłowski, Ł. (2010). Zarządzanie wynikiem finansowym w bankach z Europy Środkowo-Wschodniej związane z progowymi wartościami rentowności. *Master of Business Administration*, 5, 25–46.
- Jackowicz, K., Kuryłek, W. (2005). Unikanie raportowania strat przez banki komercyjne działające w Polsce. *Studia i Prace Kolegium Zarządzania i Finansów*, (64), 63–84.
- Jiang, F., Zhu, B., Huang, J. (2013). CEO's financial experience and earnings management. *Journal of Multinational Financial Management*, 23(3), 134–145.
- Kałośki, M., Jewartowski, T. (2017). Agency costs of overvalued equity and earnings management in companies listed on WSE. *Economics and Business Review*, 3(1), 7–37.
- Kałośki, M., Jewartowski, T. (2018). Kształtowanie zysków w spółkach notowanych na GPW. Czy charakterystyka inwestorów instytucjonalnych ma znaczenie? *Finanse, Rynki Finansowe, Ubezpieczenia*, 2(92), 255–267.
- Kałośki, M., Jewartowski, T., Mizerka, J. (in press). Capital market pressure, real earnings management, and institutional ownership stability – Evidence from Poland. *International Review of Financial Analysis*. DOI: 10.1016/j.irfa.2019.01.009.
- Lizińska, J. (2015). Manipulowanie zyskami na rynkach wschodzących i rozwiniętych – konsekwencje dla długookresowej wartości przedsiębiorstwa po dokonaniu pierwotnej emisji akcji. *Zeszyty Naukowe Uniwersytetu Szczecińskiego. Finanse, Rynki Finansowe, Ubezpieczenia*, (75), 297–308.
- Lizińska, J., Czapiewski, L. (2016). Manipulowanie zyskami przez spółki debiutujące na GPW. *Ruch Prawniczy, Ekonomiczny i Socjologiczny*, 78(4), 197–212.
- Magnis, C., Iatridis, G. E. (2017). The relation between auditor reputation, earnings and capital management in the banking sector: An international investigation. *Research in International Business and Finance*, 39, 338–357.
- Mellado-Cid, C., Jory, S. R., Ngo, T. N. (2018). Real activities manipulation and firm valuation. *Review of Quantitative Finance and Accounting*, 50(4), 1201–1226.
- Panek, T. (2009). *Statystyczne metody wielowymiarowej analizy porównawczej*. Warszawa: Szkoła Główna Handlowa w Warszawie.
- Piosik, A. (2013). *Kształtowanie zysków podmiotów sprawozdawczych w Polsce. MSR/MSSF a ustawa o rachunkowości*. Warszawa: Wydawnictwo C.H. Beck.
- Park, Y. W., Shin, H. H. (2004). Board composition and earnings management in Canada. *Journal of Corporate Finance*, 10(3), 431–457.
- Roychowdhury, S. (2006). Earnings management through real activities manipulation. *Journal of Accounting and Economics*, 42(3), 335–370.
- Shan, Y., Taylor, S. L., Walter, T. S. (2010). Errors in estimating unexpected accruals in the presence of large changes in net external financing. *SSRN Electronic Journal*. DOI: 10.2139/ssrn.1710586.
- Wójtowicz, P. (2007a). Wpływ rewizji finansowej na kształtowanie zysków w celu unikania ujawniania strat. *Zeszyty Teoretyczne Rachunkowości*, 36(92), 130–146.

Wójtowicz, P. (2007b). Kształtowanie zysków w celu unikania strat w spółkach niepublicznych. *Zeszyty Teoretyczne Rachunkowości*, 38(94), 249–264.

Wójtowicz, P. (2012). Próba identyfikacji czynników determinujących kształtowanie wyniku finansowego w celu unikania małych strat w spółkach publicznych w Polsce. *Zeszyty Teoretyczne Rachunkowości*, 65(121), 111–129.

Wójtowicz, P. (2015). Earnings management to achieve positive earnings surprises in case of medium size companies listed in Poland. *International Journal of Accounting and Economics Studies*, 3(2), 141–147.

Zang, A. Y. (2011). Evidence on the trade-off between real activities manipulation and accrual-based earnings management. *The Accounting Review*, 87(2), 675–703.

## Starzenie się ludności a redystrybucja dochodów

Andrzej Ochocki<sup>a</sup> 

**Streszczenie.** Celem artykułu jest wskazanie makroekonomicznych uwarunkowań dynamiki wynagrodzeń i emerytur w Polsce w latach 2010–2017 wraz z prognozą studialną do 2025 r. Źródłem danych są rachunki narodowe, statystyka ludności, statystyka świadczeń emerytalnych i prognoza demograficzna Głównego Urzędu Statystycznego na lata 2008–2035. Demograficzne starzenie się społeczeństwa powoduje narastanie problemu podziału bieżącego strumienia dóbr i usług wytwarzanych w procesie gospodarczym. Możliwości w tym zakresie są zdeterminowane relacjami dynamiki liczebności pracujących i emerytów oraz dynamiki wolumenu dochodu narodowego. Wynikające stąd zależności można opisać za pomocą autorskiego modelu międzypokoleniowej redystrybucji dochodów (MMRD).

Tendencje obserwowane w Polsce dowiodły, że dynamika wzrostu wolumenu PKB w badanym okresie zapewniała makroekonomiczny warunek równowagi w procesie kształtowania się wynagrodzeń i emerytur – wzrost PKB był większy niż wynagrodzeń i emerytur. Wykonane szacunki wskazują zarazem na konieczność utrzymania wysokiego tempa wzrostu PKB i produktywności na poziomie ok. 3,6% rocznie, co oznacza wzrost wolumenu PKB o ponad 40% w 2025 r. w porównaniu z rokiem 2015.

Przedstawiony model międzypokoleniowej redystrybucji dochodów pozwolił także wskazać sedno problemu publicznych transferów międzypokoleniowych, które w państwie traktowanym jako dobro wspólne należy pojmować jako wierzytelności emerytów, a ich spłacanie powinno zapewniać godziwy poziom życia ludziom starszym.

**Słowa kluczowe:** pracujący, emeryci, wynagrodzenia, emerytury, PKB, produktywność, model międzypokoleniowej redystrybucji dochodów

### Population ageing and income redistribution

**Abstract.** The aim of the paper is to identify the macroeconomic conditions which determined the dynamics of remuneration and pension growth in Poland in 2010–2017, along with presenting the study forecast until 2025. The research was based on the data from the national accounts, population statistics, statistics of retirement benefits and Statistics Poland's demographic forecast for 2008–2035. The ageing of a population aggravates the problem of fair distribution of the current stream of goods and services produced in the economic process. Distribution options are determined by the relations between the ratio of employed persons to pensioners and the volume of national income. The resulting dependencies can be described using the author's original Intergenerational Income Transfer Model (IITM).

The observation of tendencies in Poland demonstrated that the pace of GDP growth between 2010–2017 ensured the macroeconomic balance essential in the process of shaping remunerations and pensions – GDP was growing faster than remunerations and pensions. This indicates the necessity of maintaining a high rate of GDP growth and productivity, at around 3.6% a year, which means that GDP will need to have grown by over 40% by 2025 compared to 2015.

With the help of the IITM, it was also possible to identify the main difficulty facing the system of public intergenerational transfers, which, in a state perceived as a common good, should be understood as pensioners' liabilities whose repayment guarantees a decent standard of living for the elderly.

**Keywords:** employed persons, pensioners, remunerations, pensions, GDP, productivity, Intergenerational Income Transfer Model

**JEL:** J11, E24

---

<sup>a</sup> Uniwersytet Kardynała Stefana Wyszyńskiego w Warszawie, Wydział Społeczno-Ekonomiczny.

Demograficzne starzenie się ludności jest jednym z najważniejszych zjawisk współczesnego świata. Proces ten ma wymiar osobowy i społeczny, a na jego znaczenie wskazywał już nestor polskiej demografii Edward Rosset, pisząc: „Przed demografem, który zajmuje się problemem ludzi starych, stoją tak istotne kwestie, jak pozycja seniorów w społeczeństwie, zabezpieczenie ich potrzeb, źródła i konsekwencje starzenia się ludności, wpływ tego procesu na stosunki ekonomiczne, demograficzne, polityczne, psychologiczne itd.” (1967, s. 5).

Problem podziału dochodu narodowego w kontekście demograficznego starzenia się ludności należy rozważać na trzech płaszczyznach: makroekonomicznej, systemu emerytalnego i prywatnych transferów dochodów w sektorze gospodarstw domowych (Kurkiewicz, 2007). Starzenie się ludności skutkuje narastającymi trudnościami w podziale bieżącego strumienia dóbr i usług wytwarzanych w procesie gospodarczym (Légaré, 2006). Możliwości w tym zakresie są zdeterminowane relacjami dynamiki liczebności pracujących i emerytów oraz wolumenu dochodu narodowego. Zależności, jakie stąd wynikają, można opisać za pomocą modelu międzypokoleniowej redystrybucji dochodów (MMRD), opracowanego przez Ochockiego (2008).

Celem pracy jest wskazanie makroekonomicznych uwarunkowań dynamiki wynagrodzeń i emerytur w Polsce w latach 2010–2017 wraz z prognozą studialną do 2025 r. Źródłem danych są rachunki narodowe, statystyka ludności, statystyka świadczeń emerytalnych i prognoza demograficzna Głównego Urzędu Statystycznego (GUS).

## MODEL MIĘDZYPOKOLENIOWEJ REDYSTRYBUCJI DOCHODÓW

W podziale dochodu narodowego – w jego części przeznaczony na spożycie – kluczowe znaczenie mają proporcje między wynagrodzeniami i emeryturami oraz udział wynagrodzeń i emerytur w PKB. Proporcje te – przy założeniu *ceteris paribus* – są kształtowane przez następujące czynniki:

- produktywność zasobów pracy;
- poziom wynagradzania za pracę;
- stosunek liczebności pracujących do emerytów;
- wielkość wierzytelności emerytalnych.

Równowagę w procesie podziału dochodu narodowego w części przeznaczony na spożycie zapewnia równość relacji przeciętnego wynagrodzenia do przeciętnej kwoty wypłaconej z funduszy emerytalnych, obliczonej na jedną osobę pracującą na rynku pracy w dwóch kolejnych okresach MMRD (Ochocki, 2008):

$$\frac{\frac{\alpha(P_a \cdot L_a)}{L_a}}{\frac{L_e \cdot R}{L_a}} = \frac{\frac{\alpha'(P'_a \cdot L'_a)}{L'_a}}{\frac{L'_e \cdot R'}{L'_a}} \quad (1)$$

gdzie:

$L_a$  – liczba pracujących,

$L_e$  – liczba emerytów,

$P_a$  – przeciętna produktywność osoby zawodowo czynnej ( $PKB/L_a$ ),

$\alpha$  – relacja przeciętnego wynagrodzenia do przeciętnej produktywności osoby zawodowo czynnej (udział wynagrodzeń w PKB),

$R$  – kwota przeciętnej emerytury.

Z równości (1)<sup>1</sup> wynikają – przy założeniu stałych cen, niezmiennych proporcji podziału dochodu narodowego na spożycie i akumulację, stałego udziału wydatków publicznych i udziału wynagrodzeń w dochodzie narodowym, a jednocześnie w warunkach określonych zmian produktywności oraz liczebności pracujących i emerytów – trzy następujące wskaźniki graniczne (modelowe):

- dynamiki siły nabywczej przeciętnej emerytury (maksimum):

$$\frac{R'}{R} = \frac{P'_a}{P_a} \cdot \frac{L'_a}{L_a} \cdot \frac{L_e}{L'_e} \quad (2)$$

- dynamiki siły nabywczej przeciętnego wynagrodzenia (maksimum), która równa się dynamice produktywności w ujęciu realnym (minimum):

$$\frac{P'_a}{R} = \frac{R'}{R} \cdot \frac{L'_e}{L_e} \cdot \frac{L_a}{L'_a} \quad (3)$$

- dynamiki PKB w ujęciu realnym (minimum):

$$\frac{PKB'}{PKB} = \frac{R'}{R} \cdot \frac{L'_e}{L_e} \quad (4)$$

Przy założeniu  $\alpha = \text{constans}$  zachowanie dynamiki realnych wynagrodzeń i emerytur w ramach określonych wskaźnikami granicznymi (wzory (2), (3) i (4)) oznacza utrzymanie tempa wzrostu siły nabywczej wynagrodzeń i emerytur na poziomie nieprzekraczającym tempa wzrostu produktywności osób pracujących.

W sytuacji jednoczesnego przekroczenia granicznej dynamiki wynagrodzeń i granicznej dynamiki emerytur łączna siła nabywcza dochodów z pracy i emerytur przewyższa uprzednio ukształtowaną część wolumenu PKB uzyskiwanego przez pracujących i emerytów. Zachodzi wtedy konieczność równoważenia podziału PKB. W tym celu uruchomiony zostaje mechanizm osiągnięcia nowego stanu równowagi poprzez – występujące łącznie lub alternatywnie – następujące działania:

- ustalanie nowego poziomu wynagrodzeń za pracę;
- kształtowanie adekwatnych systemów emerytalnych;
- zmiana dynamiki inwestowania w sektorach prywatnym i publicznym gospodarki;

<sup>1</sup> Wzór ten można zredukować do postaci  $\frac{\alpha(P_a \cdot L_a)}{L_e \cdot R} = \frac{\alpha'(P'_a \cdot L'_a)}{L'_e \cdot R'}$ , gdzie  $\alpha(P_a \cdot L_a)$  i  $\alpha'(P'_a \cdot L'_a)$  – wartości wypłaconych wynagrodzeń,  $L_e \cdot R$  i  $L'_a \cdot R'$  – kwoty wypłaconych emerytur.



- modyfikowanie reguł podatkowych i wydatków publicznych państwa;
- narastanie długu publicznego;
- zmiana oszczędności i zadłużenia gospodarstw domowych;
- podwyższanie i zmiana relacji cen towarów i usług konsumpcyjnych.

Ich wynikiem są nowe relacje produktywności, wynagrodzeń i emerytur, a w konsekwencji – nowe proporcje podziału PKB na spożycie i akumulację (Branson, 1989).

Na tej podstawie możemy twierdzić, że w podziale dochodu narodowego w gospodarce rynkowej działa prawo równowagi wynagrodzeń i emerytur, które uruchamia działania dostosowawcze w razie przekroczenia granicznych wskaźników dynamiki tych dwóch parametrów, w warunkach określonej produktywności pracujących oraz liczebności populacji pracowników i emerytów.

### SYTUACJA NA RYNKU PRACY

Proces demograficznego starzenia się ludności zmniejsza proporcję osób w wieku produkcyjnym w stosunku do tych w wieku poprodukcyjnym, a w konsekwencji uszczupla zasoby pracy. W Polsce ta tendencja występuje od wielu lat; tylko w okresie od 2010 do 2017 r. opisana proporcja spadła z 3,85 do 2,94 (tabl. 1). Według prognozy ludności GUS w 2025 r. wyniesie ona 2,40 (GUS, 2009, tabl. A.8).

**TABL. 1. LUDNOŚĆ W WIEKU PRODUKCYJNYM I POPRODUKCYJNYM ORAZ PRACUJĄCY I EMERYCI**

Wyszczególnienie	2010	2015	2017
Ludność w wieku produkcyjnym $W_{pr}$ w tys. <sup>a</sup> .....	24831,0	24002,2	23517,6
Ludność w wieku poprodukcyjnym $W_p$ w tys. <sup>b</sup> .....	6456,6	7533,3	7995,3
Relacja $W_{pr} : W_p$ .....	3,85	3,19	2,94
Pracujący w gospodarce $L_a$ w tys. <sup>c</sup> .....	14865,0	15086,0	15820,0
Emeryci $L_e$ w tys. <sup>d</sup> .....	6353,0	6253,0	6435,0
Relacja $L_a : L_e$ .....	2,33	2,42	2,46
Wskaźnik zatrudnienia w % <sup>e</sup> .....	50,0	51,9	53,7
Wskaźnik bezrobocia w % <sup>f</sup> .....	9,6	7,5	4,9

a Mężczyźni 18–64 lata, kobiety 18–59 lat. b Mężczyźni 65 lat i więcej, kobiety 60 lat i więcej. c Pracujący w ujęciu rachunków narodowych, tj. osoby wykonujące pracę przynoszącą im zarobek lub dochód łącznie z szacunkową liczbą pracujących w szarej gospodarce w przeliczeniu na pełnozatrudnionych. d Ubezpieczeni w systemach społecznych nierolniczym (FUS, budżet instytucji państwowych) i rolniczym (KRUS). e Udział pracujących w populacji ludności w wieku 15 lat i więcej według BAEL. f Udział bezrobotnych w populacji ludności aktywnej zawodowo według BAEL.

Źródło: GUS (2018) i obliczenia własne.

Należy jednak podkreślić, że w latach 2010–2017 zwiększała się liczba pracujących w całej gospodarce na skutek wzrostu aktywności zawodowej ludności, a także napływu pracowników z zagranicy, zwłaszcza z Ukrainy (Górny, 2017). Jednocześnie zwiększała się liczba emerytów, ale w wolniejszym tempie, dzięki

czemu relacja liczby pracujących w stosunku do liczby emerytów poprawiła się i w 2017 r. wyniosła 2,33 (w 2010 r. było to 2,46).

## UDZIAŁ WYNAGRODZEŃ I EMERYTUR W WARTOŚCI PKB

Udział nominalnych wynagrodzeń, jak również emerytur w PKB nie uległ dużym zmianom w latach 2010–2017. Zwiększył się natomiast wskaźnik relacji realnej kwoty wypłaconych emerytur do liczby pracujących w całej gospodarce (tabl. 2).

**TABL. 2. RZECZYWISTY UDZIAŁ WYNAGRODZEŃ I EMERYTUR W PKB ORAZ DYNAMIKA WSKAŹNIKA RELACJI ŁĄCZNEJ KWOTY ŚWIADCZEŃ EMERYTALNYCH DO LICZBY PRACUJĄCYCH W GOSPODARCE**

Wyszczególnienie	2010	2015	2017
Udział nominalnych wynagrodzeń w PKB <sup>a</sup> w %	32,6	31,1	31,8
Udział nominalnych emerytur w PKB <sup>b</sup> w %.....	8,6	8,4	8,2
Dynamika wskaźnika relacji realnej kwoty emerytur do liczby pracujących <sup>c</sup> .....	100,0	111,6	112,4

a Wynagrodzenia i dochody z pracy brutto łącznie z rolnictwem w ujęciu rachunków narodowych (GUS, 2018).  
b Kwota brutto świadczeń emerytalnych z ubezpieczeń społecznych w systemach nierolniczym (FUS, budżet instytucji państwowych) i rolniczym (KRUS). c Siła nabywcza łącznej kwoty świadczeń emerytalnych w cenach z 2010 r.

Ź r ó d ł o: jak przy tabl. 1.

Wzrost wskaźnika realnej kwoty wypłaconych emerytur przypadającej na osobę pracującą w całej gospodarce odzwierciedla potrzebę wzrostu wolumenu PKB na skutek procesu demograficznego starzenia się ludności i zwiększenia się siły nabywczej przeciętnej emerytury.

## PKB, PRODUKTYWNOŚĆ ORAZ PRZECIĘTNA WARTOŚĆ WYNAGRODZENIA I EMERYTURY

Dynamikę monetarnych parametrów uwzględnionych w MMRD przedstawia tabl. 3.

**TABL. 3. DYNAMIKA WOLUMENU PKB, PRODUKTYWNOŚCI OSOBY PRACUJĄCEJ, PRZECIĘTNEGO WYNAGRODZENIA REALNEGO W GOSPODARCE I PRZECIĘTNEJ REALNEJ EMERYTURY (ceny stałe)**

Wyszczególnienie		2015	2017	
		2010 = 100	2015 = 100	
Wolumen PKB .....	a	116,0	125,2	107,9
	b	113,6	119,8	105,5
Przeciętna produktywność w gospodarce <sup>2</sup>	a	114,3	117,7	103,0
	b	111,9	112,6	100,6

1 Dynamikę graniczną obliczono na podstawie wzorów odpowiednio: (2), (3) i (4). 2 Iloraz wolumenu PKB i liczby pracujących w gospodarce w ujęciu rachunków narodowych.

**TABL. 3. DYNAMIKA WOLUMENU PKB, PRODUKTYWNOŚCI OSOBY PRACUJĄCEJ, PRZECIĘTNEGO WYNAGRODZENIA REALNEGO W GOSPODARCE I PRZECIĘTNEJ REALNEJ EMERYTURY (dok.)**

Wyszczególnienie a – dynamika rzeczywista b – dynamika graniczna (minimum) <sup>1</sup>		2015	2017	
		2010 = 100	2010 = 100	2015 = 100
Przeciętne wynagrodzenie realne <sup>3</sup> .....	a	108,3	115,1	106,3
	b	111,9	112,6	100,6
Przeciętna emerytura realna <sup>4</sup> .....	a	115,4	118,3	102,5
	b	117,9	123,7	104,9

<sup>1</sup> Dynamikę graniczną obliczono na podstawie wzorów odpowiednio: (2), (3) i (4). <sup>3</sup> Iloraz realnych wynagrodzeń i dochodów z pracy (w tym w rolnictwie indywidualnym) brutto w stosunku do liczby pracujących w gospodarce w ujęciu rachunków narodowych. <sup>4</sup> Łącznie z systemów ubezpieczeń społecznych pozarolniczego i rolniczego.

Źródło: GUS (2018), Ochocki (2012) i obliczenia własne.

Z analizy rzeczywistych i granicznych wskaźników dynamiki wartości parametrów monetarnych, uwzględnionych w MMRD wynika, że:

- rzeczywista dynamika wolumenu PKB i produktywności pracujących była wyższa niż modelowa w latach 2010–2017;
- rzeczywista dynamika produktywności była niższa niż rzeczywista dynamika wolumenu PKB w latach 2010–2017;
- rzeczywista dynamika siły nabywczej przeciętne go wynagrodzenia w gospodarce nie przekraczała dynamiki granicznej w latach 2010–2015, natomiast była wyższa niż graniczna w latach 2016 i 2017;
- rzeczywista dynamika siły nabywczej przeciętne go wynagrodzenia była niższa niż rzeczywista dynamika produktywności pracujących w latach 2010–2015, natomiast odwrotne tendencje zaobserwowano w latach 2016 i 2017, kiedy dynamika produktywności była dwukrotnie niższa niż dynamika siły nabywczej przeciętne go wynagrodzenia;
- rzeczywista dynamika siły nabywczej przeciętne j emerytury nie przekraczała dynamiki granicznej w latach 2010–2017.

Zaobserwowane tendencje dowodzą, że dynamika wzrostu wolumenu PKB w całym badanym okresie 2010–2017 zapewniała makroekonomiczny warunek równowagi w procesie kształtowania się wynagrodzeń i emerytur. Dynamika wzrostu PKB przewyższała dynamikę wzrostu przeciętne go wynagrodzenia i przeciętne j emerytury. Jednakże na wzrost gospodarczy w tym okresie bardziej wpłynęło zwiększenie się liczby pracujących niż produktywności. Jednocześnie realne wynagrodzenia rosły szybciej niż produktywność pracujących w gospodarce w latach 2016 i 2017, co przy ograniczonych możliwościach zwiększania zatrudnienia w następnych latach może uniemożliwić utrzymanie na dotychczasowym poziomie realnej siły nabywczej emerytur, zwłaszcza w warunkach nasilonego starzenia się ludności.

## WZROST GOSPODARCZY

Jak wykazano, MMRD uwidacznia uwarunkowania makroekonomiczne i mechanizm kształtowania się wynagrodzeń i emerytur w podziale dochodu narodowego. Z tego modelu wynika, że proces starzenia się ludności określa minimalny wzrost PKB, który umożliwi utrzymanie siły nabywczej emerytur i wynagrodzeń, przy założeniu stałego udziału wynagrodzeń w PKB oraz rzeczywistej dynamiki wzrostu liczby emerytów.

Przyjmując założenia prognozy studialnej, że:

- relacja liczby ludności w wieku produkcyjnym w stosunku do liczby ludności w wieku poprodukcyjnym zmniejszy się do 2,40 w 2025 r. (GUS, 2009);
- stopa bezrobocia zostanie utrzymana na poziomie ok. 5%;
- liczba pracujących nie ulegnie zmianie i wyniesie 15086 tys. osób w 2025 r.;
- liczba emerytów wyniesie 8097 tys. osób w 2025 r. (Ochocki, 2012);
- relacja liczby pracujących i emerytów wyniesie 1,86 w 2025 r.;
- siła nabywcza przeciętnej wynagrodzenia będzie wzrastać w tempie wzrostu przeciętnej produktywności osoby pracującej;
- siła nabywcza przeciętnej emerytury będzie wzrastać o 1% rocznie;
- $\alpha = \text{constans}$

można oszacować minimalne tempo wzrostu PKB i produktywności pracujących w latach 2016–2025, które wyniesie 3,6% rocznie (tabl. 4).

**TABL. 4. DYNAMIKA WOLUMENU PKB I PRODUKTYWNOŚCI OSOBY PRACUJĄCEJ**  
(ceny stałe)

Wyszczególnienie	2011–2015 <sup>a</sup>	2016 i 2017 <sup>a</sup>	2016–2025 <sup>b</sup>
	rok poprzedni = 100		
Wolumen PKB .....	103,0	103,8	103,6
Produktywność osoby pracującej w gospodarce	102,7	101,5	103,6

a Wartości rzeczywiste. b Wartości prognozowane.

Źródło: jak przy tabl. 3.

Wykonane szacunki wskazują na konieczność utrzymania dużego tempa wzrostu PKB i produktywności na poziomie ok. 3,6% rocznie, co oznacza więcej niż podwojenie średniorocznej dynamiki produktywności do 2025 r. Wyniki te wskazują zarazem na konieczny wzrost wolumenu PKB o ponad 40% w 2025 r. w porównaniu z 2015 r. Osiąganie niższych wskaźników wzrostu gospodarczego zakłóci równowagę w procesie podziału dochodu narodowego na spożycie i akumulację oraz wywoła impuls inflacyjny.

Wysoka dynamika wzrostu PKB i relatywnie niższe tempo wzrostu produktywności byłyby możliwe w sytuacji utrzymania masowego napływu imigrantów na rynek pracy, wysokiej koniunktury gospodarczej w Europie, odpowiedniej stopy inwestowania oraz możliwości zaspokojenia potrzeb mieszkaniowych imigrantów podejmujących pracę w Polsce. Autorska prognoza ludności do 2060 r. (Anacka i Janicka, 2018) wskazuje na wysoką stopę imigracji w Polsce.

## KONSENSUS SPOŁECZNY

Proces starzenia się ludności kształtuje realne wyzwania zarówno dla gospodarki, jak i dla całego społeczeństwa (Schoenmaeckers i Vanderleyden, 2005). Demograficzne starzenie się ludności – jak wynika z dotychczasowych doświadczeń krajów europejskich – powoduje wzrost udziału wydatków na emerytury w PKB bez względu na rozwiązania systemowe w dziedzinie ubezpieczeń społecznych. Narasta więc problem podziału dochodów realnych w relacjach między-pokoleniowych.

Dynamika produktywności i potrzeby rozwojowe gospodarki wyznaczają płaszczyznę możliwego konsensusu społecznego w sprawie podziału dochodów między pracownikami i emerytami. W starzejącym się społeczeństwie osiągnięcie takiego konsensusu jest trudne nawet w warunkach odpowiedniego wzrostu gospodarczego, ponieważ istotną rolę odgrywają tu co najmniej trzy grupy czynników: ekonomiczne, socjologiczne i psychologiczne. Wzrost produktywności wiąże się zwykle z procesami modernizacyjnymi w gospodarce, co wymaga nie tylko zwiększenia inwestycji, lecz także zmiany struktury zawodowej ludności, która w warunkach starzenia się ludności napotyka wiele barier różnorodnej natury. Z kolei zwiększony wysiłek społeczny nakierowany na wzrost produktywności bez wzrostu realnych dochodów pracowników wywołuje zwykle sprzeciw związków zawodowych i powoduje napięcia społeczne (Piketty, 2015). Na płaszczyźnie psychologicznej stagnacja dochodów pracowników osłabia motywacyjną funkcję wynagrodzeń. Próby dokonywania zmian transferów społecznych, polegające na ich uszczupleniu dla dzieci i młodzieży i jednoczesnym powiększaniu tych transferów w celu utrzymania poziomu realnej emerytury, są praktycznie niewykonalne chociażby ze względu na rosnące aspiracje społeczne i koszty edukacji, a także rosnące wydatki na ochronę zdrowia ludzi w starszym wieku.

Dyskusja i spory wokół problemu odzyskiwania wiarygodności emerytalnych w warunkach starzenia się ludności mogą więc prowadzić do konfliktów pokoleniowych, a w konsekwencji osłabiać więzi społeczne. W związku z tą kwestią wybitny demograf Alfred Sauvy konstataował, że prawdopodobnie możliwe będzie przystosowanie się społeczeństw do starości demograficznej, a nawet zmniejszania się liczebności populacji. Warunkiem takiego przystosowania musi być jednak odpowiednie ukształtowanie świadomości społecznej (Sauvy, 1963).

## PODSUMOWANIE

Demograficzne starzenie się ludności powoduje narastanie problemu podziału bieżącego strumienia dóbr i usług wytwarzanych w procesie gospodarczym. Możliwości w tym zakresie są zdeterminowane relacjami dynamiki liczebności pracujących i emerytów oraz dynamiki wolumenu dochodu narodowego. Analiza tych procesów obserwowanych w Polsce dowiodła, że dynamika wzrostu wolu-

menu PKB w całym badanym okresie zapewniała makroekonomiczny warunek równowagi w procesie kształtowania się dochodów z pracy i emerytur: wolumen PKB wzrastał szybciej niż siła nabywcza dochodów z pracy i emerytur. Podobne relacje występowały w latach 2016 i 2017. Jednocześnie w tych dwóch latach nastąpiło wyraźne przyspieszenie wzrostu siły nabywczej dochodów z pracy, których realna przeciętna wartość rosta szybciej niż przeciętna produktywność osoby zawodowo czynnej. Prognoza studialna, zakładająca udział wartości wynagrodzeń i innych dochodów z pracy w PKB na obecnym poziomie i niewielki wzrost siły nabywczej przeciętnej emerytury o 1% rocznie do 2025 r., wykazała konieczność utrzymania wysokiego tempa wzrostu PKB i produktywności na poziomie ok. 3,6% rocznie, co oznacza wzrost wolumenu PKB o ponad 40% w 2025 r. w porównaniu z rokiem 2015.

Dynamika PKB i potrzeby rozwojowe gospodarki wyznaczają płaszczyznę możliwego konsensusu społecznego w sprawie podziału dochodów między pracującymi a emerytami. Kluczowymi przesłankami w osiągnięciu takiego konsensusu są wzrost produktywności osób pracujących i dialog społeczny w sprawie transferów socjalnych oraz skali redystrybucji dochodów ludności. W dialogu tym niezbędne jest odwoływanie się do aksjologicznych aspektów procesu starzenia się ludności w kontekście dochodów gospodarstw domowych pracujących i emerytów.

Przedstawiony model międzypokoleniowej redystrybucji dochodów pozwolił wskazać znaczenie publicznych transferów międzypokoleniowych w procesie podziału dochodu narodowego, którego skalę odzwierciedla relacja dochodów z pracy do wypłacanych emerytur. Fundusze emerytalne w państwie traktowanym jako dobro wspólne należy pojmować jako wierzytelności emerytów, a ich spłacanie powinno zapewniać godziwy poziom życia ludziom starszym.

## BIBLIOGRAFIA

- Anacka, M., Janicka, A. (2018). Prognoza ludności Polski na podstawie ekonometrycznej prognozy strumieni migracyjnych. *Wiadomości Statystyczne*, (8), 5–27.
- Branson, W. H. (1989). *Macroeconomic. Theory and Policy*. Third Edition. New York: Harper & Row.
- Górny, A. (2017). All circular but different. Variation in patterns of Ukraine-to-Poland migration. *Population, Space and Place*, 23(8), 1–10. DOI: 10.1002/psp.2074.
- GUS. (2009). *Prognoza ludności na lata 2008–2035*. Warszawa: Główny Urząd Statystyczny.
- GUS. (2018). *Rocznik Statystyczny 2018*. Warszawa: Główny Urząd Statystyczny.
- Kurkiewicz, J. (red.). (2007). *Ludzie starsi w rodzinie i społeczeństwie*. Kraków: Wydawnictwo Uniwersytetu Ekonomicznego.
- Légaré, J. (2006). Economic, Social, and Cultural Consequences of the Aging of the Population. W: G. Caselli, J. Vallin, G. Wunsch (red.), *Demography: Analysis and synthesis*, 3 (s. 327–336). London, New York, Tokyo: Elsevier.

- Ochocki, A. (2008). Intergenerational Income Transfer Model. *Statistics in Transition new series*, 9(3), 561–570.
- Ochocki, A. (2012). Redystrybucja dochodów w perspektywie starzenia się ludności Polski. W: A. Rączaszek (red.), *Demograficzne uwarunkowania rozwoju gospodarczego*, Studia Ekonomiczne, Zeszyty Naukowe nr 103 (s. 9–16). Katowice: Wydawnictwo Uniwersytetu Ekonomicznego.
- Piketty, T. (2015). *Ekonomia nierówności*. Warszawa: Wydawnictwo Krytyki Politycznej.
- Rosset, E. (1967). *Ludzie starzy. Studium demograficzne*. Warszawa: Państwowe Wydawnictwo Ekonomiczne.
- Sauvy, A. (1963). *Théorie générale de la population*, vol. I i II. Paris: Presses Universitaires de France.
- Schoenmaeckers, R. C., Vanderleyden, L. (2005). Intergenerational Solidarity, the Elderly and Ageing: Main Results. *Studia Demograficzne*, 2(148), 100–113.

## Spis ludności na ziemiach wschodnich Rzeczypospolitej Polskiej w 1919 r.

Bożena Łazowska<sup>a</sup> 

**Streszczenie.** Celem artykułu jest przedstawienie organizacji oraz wyników spisu ludności przeprowadzonego w grudniu 1919 r. na ziemiach wschodnich Rzeczypospolitej Polskiej (RP), administrowanych przez Zarząd Cywilny Ziem Wschodnich (ZCZW). Głównym zadaniem ZCZW było sporządzenie bilansu strat wojennych i zbudowanie zrębów polskiej administracji na terenach Litwy historycznej, które weszły w skład odrodzonego państwa polskiego. Temu właśnie służył spis ludności w 1919 r. Jego wyniki miały duże znaczenie zarówno w historii statystyki, jak i historii Polski. Zostały w znacznej części włączone do pierwszego powszechnego spisu ludności RP przeprowadzonego przez Główny Urząd Statystyczny we wrześniu 1921 r. Umożliwiły opracowanie tymczasowego podziału administracyjnego ziem wcielonych do RP w 1921 r. po podpisaniu pokoju z Rosją. Korzystano z nich także przy tworzeniu planu rozwoju ziem wschodnich w odrodzonej Polsce.

**Słowa kluczowe:** historia statystyki, spis ludności, Zarząd Cywilny Ziem Wschodnich

## Census of the Eastern Lands of the Second Polish Republic in 1919

**Abstract.** The purpose of the paper is to present the process of organization and the results of the census carried out in December 1919 in the Eastern Lands of the Second Polish Republic by the Civil Administration of the Eastern Lands. The main tasks of the Civil Administration of the Eastern Lands included the assessment of the size of war losses as well as building the foundations for the Polish administration system in those territories of historical Lithuania that had become a part of the reborn Polish state. Hence the census in 1919. Its results have been of great importance for both the history of statistics and the history of Poland. In the main part, they were incorporated into the first general census of the Polish population carried out by Statistics Poland (GUS) in September 1921. Additionally, those data made it possible to develop a temporary administrative division of the territories incorporated into the Second Polish Republic in 1921, after signing a peace treaty with Russia. They were also used while forming a development plan for the eastern territories of the reborn Poland.

**Keywords:** history of statistics, census, Civilian Administration of the Eastern Lands

**JEL:** N0, N9

---

<sup>a</sup> Centralna Biblioteka Statystyczna im. Stefana Szulca.



Po zakończeniu I wojny światowej i odzyskaniu przez Polskę niepodległości władzę na ziemiach wschodnich, początkowo sprawowaną przez Naczelnego Dowódcę Wojsk Polskich Józefa Piłsudskiego, objął Zarząd Cywilny Ziem Wschodnich (ZCZW). Został on utworzony dekretem z 19 lutego 1919 r. przez Generalnego Komisarza Cywilnego. Z rozkazu Piłsudskiego bezpośredni nadzór nad ZCZW objął rząd Rzeczypospolitej Polskiej (RP) z premierem Ignacym Janem Paderewskim na czele. Od 19 lutego do 5 maja 1919 r. Komisarzem Cywilnym Ziem Wschodnich był Ludwik Kolankowski – historyk i działacz polityczny (Majchrowski, 1994). Następnie rozkazem Naczelnego Wodza Wojsk Polskich Józefa Piłsudskiego z 15 kwietnia 1919 r. Komisarzem Generalnym Ziem Wschodnich mianowano Jerzego Osmołowskiego – działacza społecznego i politycznego (Majchrowski, 1994).

W kolejnych miesiącach pod administracją ZCZW znalazły się:

- powiat miejski w Wilnie, utworzony 28 maja 1919 r.;
- okręg brzeski, utworzony 7 czerwca 1919 r., obejmujący początkowo powiaty: brzeski, wołkowyski, prużański, słonimski, kobryński i piński, do których przyłączono 6 listopada 1919 r. powiat mozyrski;
- okręg wileński, utworzony również 7 czerwca 1919 r., w którego skład weszły początkowo powiaty: wileński, trocki, oszmiański, święciański, lidzki, grodzieński i nowogródzki, od 23 października 1919 r. – powiat brasławski, a od 6 listopada 1919 r. – powiat dziśnieński;
- okręg wołyński, utworzony 9 września 1919 r. w miejsce zarządu powiatów wołyńskich, złożonego z powiatów: włodzimierskiego, kowelskiego i łuckiego, do których wraz z utworzeniem okręgu dołączyły: dubieński, rówieński, krzemieniecki, ostrogski, zasławski i zwiahelski;
- okręg miński, utworzony 15 września 1919 r., złożony z powiatów: mińskiego, słuckiego, bobrujskiego, borysowskiego i ihumeńskiego.

Zarząd Cywilny Ziem Wschodnich w zasadzie przestał istnieć z chwilą zajęcia tych ziem przez bolszewików latem 1920 r., a formalnie został zlikwidowany 9 września 1920 r. (Dominiczak, 1992; Nowak, 1993).

## OKOLICZNOŚCI SPISU W 1919 R.

Głównym zadaniem ZCZW było przygotowanie podwalin odbudowy państwa polskiego na terenach przez niego administrowanych. Aby to umożliwić, należało w pierwszej kolejności poznać dokładną strukturę ludności zamieszkującej te ziemie – wcześniej nieznaną, o czym świadczy np. publikacja Stanisława Thugutta (1915). Dlatego w grudniu 1919 r. przeprowadzono tam spis ludności. Realizację tego zadania powierzono Centralnemu Biuru Spisu Ludności (CBSL) w Wilnie pod kierownictwem naczelnika Wydziału ZCZW, członka Polskiej Or-

ganizacji Wojskowej (POW) mjr. Zygmunta Zabierzowskiego. Komisarzami spisowymi często byli członkowie Towarzystwa Straży Kresowej i urzędnicy szczebla powiatowego (Siemakowicz, 1998).

Spis w 1919 r. był pierwszym spisem ludności przeprowadzonym przez polską administrację cywilną na ziemiach wschodnich od 1789 r. (Rzepakowski, 2005). Dla przypomnienia, Pierwszy Powszechny Spis Ludności odbył się dopiero we wrześniu 1921 r. Włączono do niego część wyników spisu z 1919 r., np. dane dotyczące ziemi wileńskiej (obszaru byłej Litwy Środkowej), której nie objął spis z 1921 r. Choć w pierwszej połowie lat 20. XX w. rząd RP planował przeprowadzenie dodatkowego spisu na Wileńszczyźnie, co znalazło wyraz w uchwałach sejmiku, to jednak ze względu na brak środków spis ten się nie odbył (Siemakowicz, 1998).

Trudno w związku z tym przecenić znaczenie spisu z 1919 r. dla formującej się administracji cywilnej państwa polskiego. Tymczasem w opublikowanych w ostatnim dwudziestolecu pracach monograficznych poświęconych ZCZW pióra Joanny Gierałtowskiej-Kałużur i Adama Janusza Mielcarka nie omawia się tego wydarzenia (Gierałtowska-Kałużur, 2002, 2003; Mielcarek, 2008, 2011). Wzmianki o spisie próżno szukać również w syntezie omawiającej historię spisów ludności Rzeczypospolitej (Strzelecki i Toczyński, 2002) oraz w monografii poświęconej tematyce i organizacji spisów powszechnych w Polsce (Olszewska i Dłuska, 1981) czy w artykule *Spisy ludności na ziemiach polskich w latach 1789–1939* ogłoszonym na łamach „Przeglądu Nauk Historycznych” (Rzepakowski, 2005). Może to wynikać z tego, że rękopiśmienne źródła dotyczące spisu zostały dopiero niedawno zdigitalizowane i udostępnione szerszemu gronu badaczy (CBSL, 1920a, 1920b) w bibliotece cyfrowej Centralnej Biblioteki Statystycznej<sup>1</sup>.

Spis ludności w 1919 r. zorganizowany na obszarze Litwy i Białorusi należącym do RP jest cennym świadectwem rodzącej się administracji Polski. Co prawda zarówno niemiecka, jak i austriacka administracja okupacyjna w okresie I wojny światowej przeprowadzały spisy ludności na mniej lub bardziej rozległych wschodnich obszarach dawniej wchodzących w skład RP, niemniej jednak spisy te nie doczekały się oficjalnej publikacji szczegółowej i według Eugeniusza Romera (1920) nie mają obiektywnej wartości naukowej.

W tajnym raporcie Erwina von Beckeratha, przedstawiciela niemieckiego sztabu generalnego w Wilnie, przesłanym 3 stycznia 1917 r. do Niemiec, czytamy:

Nie doceniano w Berlinie nie tylko liczebnej siły polskości, ale też politycznego i ekonomicznego znaczenia tego elementu. Statystyki oficjalne rosyjskie [...] zupełnie nie odpowiadają naukowym wymogom takiej pracy, są na pewno sfałszowane na niekorzyść Polaków [...]. Nasz spis z r. 1916 udowodnił, że znaczenie Polaków jest na

<sup>1</sup> <http://cbs.stat.gov.pl>.

Litwie nieporównanie potężniejsze niż przypuszczano [...]. Polacy są tam jedynym elementem posiadającym wybitne kwalifikacje polityczne i twórcze (za: Romer, 1920).

Ze spisów 1916 r. i 1918 r. przeprowadzonych przez okupacyjną administrację niemiecką na obszarze historycznej Litwy znane są tylko liczby ogółem i procentowy udział poszczególnych narodowości według powiatów. Spisów tych dokonano w różnych podziałach administracyjnych, dlatego nie można ich porównać ani między sobą, ani ze spisem rosyjskim z 1897 r. Dla przykładu spis niemiecki z 1916 r. podawał, że w powiecie wileńskim liczba mieszkańców wynosiła 197881 osób (z czego 74,3% stanowili Polacy), natomiast według spisu z 1918 r. powiat wileński zamieszkiwało 252150 osób (w tym 47,7% Polaków). Ogólne wyniki spisów niemieckich z lat 1916 i 1918 na obszarze historycznej Litwy ogłoszone zostały w Paryżu w 1919 r. przez delegację polską na konferencji pokojowej (Czekanowski, 1919). Zagadnienie spisów niemieckich z okresu I wojny światowej na terenach Królestwa Polskiego omówiono w 1919 r. na łamach „Jahrbücher für Nationalökonomie und Statistik” (Praesent, 1919). Jednak dopiero polski spis ludności z 1919 r. dał obraz całokształtu stosunków demograficznych na ziemiach wschodnich.

Obszar objęty spisem był większy niż terytorium dawnego Królestwa Polskiego, przy czym gęstość zaludnienia tych terenów należała do najniższych w Europie, często poniżej 10 osób na km<sup>2</sup>. Prawie 4 mln ludzi mieszkających w blisko 40 tys. osad używało różnych języków, wyznawało różne religie i należało do odmiennych kręgów kulturowych (Eberhardt, 1999). Ziemie wschodnie miały charakter peryferyjny i w przeważającej części były niedostępne ze względu na rozległe obszary bagienne i leśne oraz brak dróg bitych (Koter, 1997).

Polscy komisarze spisowi musieli też zmierzyć się z wielką i zrozumiałą nieufnością ludności, która pamiętając niedawne spisy niemieckie z lat 1916 i 1918, obawiała się, że prace spisowe mogłyby stać się podstawą do dokonywania grabieży majątku, podnoszenia podatków lub akcji politycznych przeciwko mieszkańcom tych ziem (Romer, 1920). Niektórzy badacze twierdzą również, że miejscowa ludność pamiętała z nie najlepszej strony także spis rosyjski przeprowadzony w 1897 r. (Kozyra, 2005).

Ziemie wschodnie były zacofane gospodarczo i cywilizacyjnie w stosunku do ziem centralnej Polski, a do tego poniosły ogromne straty podczas wojny, szczególnie dotkliwe w pasie przewlekłych działań przyfrontowych. Dokuczliwy był brak komunikacji i łączności, aprowizacji i lekarstw. W wielu powiatach widoczne było zmęczenie, a nawet demoralizacja społeczeństwa polegająca m.in. na rabunku żywności i medykamentów (Gierałtowska-Kałuża, 2003). Front I wojny światowej przez dwa lata przechodził przez te ziemie 40-kilometrowym pasem. Na samej Wileńszczyźnie zburzono blisko 87 tys. zabudowań. Niemcy wywieźli z tych terenów obiekty przemysłowe, materiały budowlane, inwentarz

żywy, dzieła sztuki, a nawet przedmioty codziennego użytku. Masowo wycinali drzewa – tylko w 1917 r. wycięli i wywieźli z rejonu Wilii więcej drewna niż w ciągu poprzednich 20 lat (Kirwiel, 2006).

Komisarze spisowi pracowali na terytorium o nieuregulowanych granicach, które rozciągało się na obszarze o długości 700 km i szerokości 500 km (Romer i Szumański, 1920). Wschodnią granicę Polski w tym okresie wyznaczały zmieniające się linie frontów. Rozpoczęta wiosną 1919 r. ofensywa wojsk polskich w końcu roku oparła się o Dźwinę, Berezynę, Bobrujsk, Ptycz, Olewsk, Zviahel, Lubar, Derażnię i dalej prowadziła na południe do Dniestru (Mielcarek, 2008). Spis grudniowy w 1919 r. objął cały obszar ziem wschodnich, tj. gubernię grodzieńską (z wyłączeniem powiatów białostockiego, bielskiego i sokólskiego), dawną gubernię wileńską (z wyjątkiem skrawków powiatu trockiego) oraz znaczną część guberni mińskiej w jej części położonej na zachód od Berezyny, a więc 172 gminy ośmiu powiatów okręgu brzeskiego, 193 gminy dziesięciu powiatów okręgu wileńskiego oraz 101 gmin sześciu powiatów okręgu mińskiego znajdujących się pod administracją ZCZW (Maliszewski, 1920). Okręgi te były identyczne z dawnymi terytoriami guberni rosyjskich, co pozwalało na porównanie wyników spisu w 1919 r. ze spisami rosyjskimi.

Materiały spisu ludności przeprowadzonego w 1919 r. trafiły do CBSL w Wilnie, a po wybuchu wojny polsko-bolszewickiej przeniesiono je do Torunia i Włocławka, z wyjątkiem manuskryptowych map i diagramów, które przewieziono do Warszawy<sup>2</sup>.

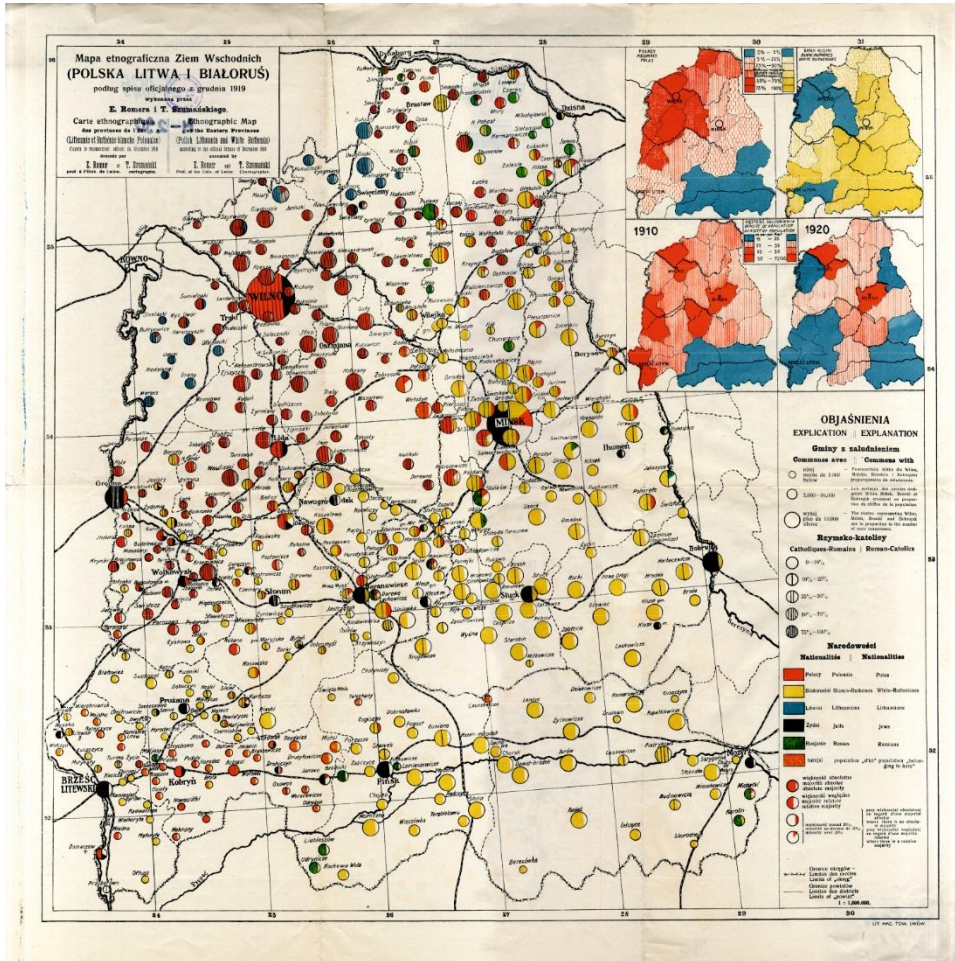
## WYNIKI SPISU

Na podstawie spisu ziem wschodnich z grudnia 1919 r. Romer opracował tablice statystyczne przedstawiające główne zagadnienia demograficzne, takie jak liczba i gęstość zaludnienia oraz stosunki wyznaniowe i narodowościowe. Była to pierwsza źródłowa i analityczna publikacja przedstawiająca wyniki tego spisu (Romer, 1920). Następnie Romer i Teofil Szumański opublikowali *Mapę etnograficzną Ziem Wschodnich* (Romer i Szumański, 1920), a Tadeusz Szturm de Sztrem i Włodzimierz Wakar wydali mapę *Stosunki narodowościowe na terenach b. Zarządu Cywilnego Ziem Wschodnich według spisu z 1919 roku* (Szturm de Sztrem i Wakar, 1921).

---

<sup>2</sup> Wówczas do Biblioteki Głównego Urzędu Statystycznego trafiły rękopiśmienne źródła: *Grafik składu wyznaniowego narodowości [i] składu narodowościowego wyznań Zarządu Cywilnego Ziem Wschodnich* opracowany przez CBSL w Wilnie w 1920 r. oraz *Grafik wyznaniowy [i] narodowościowy w procentach: okręgu: wileńskiego, brzeskiego, mińskiego łącznie z miastami, powiatów bez miast powiatowych, miast powiatowych; terenu objętego ZCZW [Zarządem Cywilnym Ziem Wschodnich]*. Grafiki te zostały zdigitalizowane i umieszczone w bibliotece cyfrowej CBS.

## ZIEMIE WSCHODNIE RZECZYPOSPOLITEJ POLSKIEJ



Źródło: Romer i Szumański (1920).

Ogółem na obszarze Litwy historycznej mieszkało w grudniu 1919 r. 3846620 osób, w tym w okręgu wileńskim – 1633504 osoby, w brzeskim – 1121978 osób, a w mińskim – 1091138 osób. Odnotowano jedynie osiem miast powyżej 10 tys. mieszkańców, tj.: Wilno – 128954, Mińsk – 102392, Bobrujsk – 29704, Grodno – 28165, Pińsk – 21436, Brześć Litewski – 14005, Lida – 11365, Baranowicze – 10373.

Skład wyznaniowy ludności Litwy historycznej był niezwykle zróżnicowany. Na ziemiach wschodnich mieszkało: 1982893 prawosławnych (którzy stanowili 51,5% ludności), 1345620 katolików (35%), 462078 wyznawców judaizmu (12%) i 54214 wiernych innych religii (1,4%, w tym: 32452 staroobrzędowców, 6702 unitów, 8726 ewangelików, 5639 muzułmanów).

Największą grupą narodowościową zamieszkującą ziemie wschodnie byli Białorusini (1615911 osób; 43% ogółu), następnie: Polacy (1218796 osób; 32%), Żydzi (392580 osób; 10,2%), Rosjanie (220235 osób; 5,7%), tzw. tutejsi<sup>3</sup> (191981 osób; 5%), Litwini (114595 osób; 3,0%), Rusini (4129 osób; 0,13%), Tatarzy (3741 osób; 0,09%) i inni (6890 osób; 0,2%). Znamienny był fakt, że aż 41762 osoby nie podały żadnej narodowości i nie uznawały się za tzw. tutejszych. Można przypuszczać, że obawiały się podać swoją narodowość nowej polskiej administracji<sup>4</sup>.

Różnice w składzie narodowościowym poszczególnych okręgów były bardzo duże. W okręgu wileńskim przeważali Polacy (874756 mieszkańców), a drugą co do wielkości grupę narodowościową stanowili Białorusini (350391 mieszkańców). W okręgu tym mieszkało też: 132089 Żydów, 112262 Litwinów, 102022 Rosjan, 37089 tutejszych i 1316 Tatarów, 3544 osoby podały inną narodowość niż powyższe, a 19521 osób w ogóle nie podało narodowości. W okręgu brzeskim przeważali Białorusini (598102 osoby), a Polacy stanowili drugą co do wielkości grupę narodowościową (184334 osoby). W tym okręgu mieszkało również: 136259 Żydów, 116965 tutejszych, 57830 Rosjan, 4129 Rusinów, 401 Litwinów i 381 Tatarów; 85917 mieszkańców podało inną narodowość niż powyższe, natomiast aż 21650 mieszkańców odmówiło podania narodowości. W okręgu mińskim przeważali Białorusini (703418) przed Polakami (159706). Kolejnymi grupami narodowymi w tym okręgu byli: Żydzi (123718), Rosjanie (60383), tutejsi (37927), Litwini (1932) i osoby innej narodowości (1419), przy czym jedynie 591 mieszkańców nie udzieliło odpowiedzi w tej kwestii.

O ile w większości miast okręgu wileńskiego przeważali Polacy, stanowiący 87,3% mieszkańców powiatu wileńskiego (w tym 56,0% mieszkańców miasta Wilno), 52,5% – brasławskiego, 76,0% – lidzkiego, 68,3% – oszmiańskiego i 53,8% – trockiego, o tyle zdecydowaną większość mieszkańców wielu miast okręgu brzeskiego i mińskiego stanowili Białorusini (62,1% mieszkańców powiatu baranowickiego, 78,3% – mozyrskiego, 68,1% – pińskiego, 66,6% – prużańskiego, 74,4% – bobrujskiego, 67,8% – borysowskiego, 65,7% – ihumeńskiego i 85,3% – słuckiego). Żydzi stanowili 71,1% mieszkańców Brześcia Litewskiego i 44,9% ludności Mińska, a w pozostałych powiatach ich odsetek wahał się od 4% do 12% wszystkich mieszkańców.

Skład wyznaniowy najliczniejszych grup przedstawiał się następująco: 91,0% Polaków wyznawało religię katolicką, 5,4% – prawosławną, a 3,0% – judaizm; 92,3% Białorusinów zadeklarowało wyznanie prawosławne, 6,0% – katolickie,

<sup>3</sup> Osoby podające się za „tutejszych” lub „miejskowych” miały jeszcze niewykształcone poczucie przynależności narodowej.

<sup>4</sup> Nie zachowały się ani żadne prace metodyczne czy założenia metodologiczne, ani formularze spisowe, na podstawie których można by ustalić definicję narodowości przyjętą w spisie.

a 1,0% – judaizm; 97,8% Litwinów wskazało katolicyzm, 1,3% – prawosławie, a 0,6% – judaizm. Spośród tutejszych 92,8% zadeklarowało wyznanie prawosławne, 6,7% – katolickie, zaś 0,5% – judaizm.

Spis ludności na ziemiach wschodnich w 1919 r. pokazał olbrzymi spadek zaludnienia w wyniku I wojny światowej. W powiecie brzeskim w 1910 r. mieszkały 54 osoby na km<sup>2</sup>, a według spisu z 1919 r. już tylko 17 osób na km<sup>2</sup>, czyli gęstość zaludnienia była o ok. 70% niższa. W powiecie wołkowyskim przeciętne zaludnienie wynosiło w 1910 r. 50 osób na km<sup>2</sup>, a w roku 1919 – już tylko 25 osób na km<sup>2</sup>, czyli zmniejszyło się o 50%. W innych powiatach spadek liczby zaludnienia nie był już tak duży, ale nadal istotny – częstokroć przekraczał 30% ogólnej liczby ludności w danym powiecie sprzed I wojny światowej (Romer, 1920). Liczby te świadczą o skali strat ludności i migracji, które wynikały z prowadzonych działań wojennych (Dmitruk, 2017; Holzer i Molenda, 1963).

**TABL. 1. LUDNOŚĆ NA ZIEMIACH WSCHODNICH W GRUDNIU 1919 R.**

Okręgi i powiaty	Powierzchnia w km <sup>2</sup>	Liczba ludności	Gęstość zaludnienia na km <sup>2</sup>		Gminy	Osady
			1919	1910		
<b>O g ó ł e m</b> .....	<b>142171</b>	<b>3846620</b>	<b>27</b>	<b>39</b>	<b>479</b>	<b>37831</b>
<b>Wileński</b> .....	<b>48466</b>	<b>1633504</b>	<b>34</b>	<b>48</b>	<b>205</b>	<b>23490</b>
Brasławski .....	2590	82513	32	44	15	1901
Dziśnieński .....	5779	193263	33	43	24	3118
Grodzieński <sup>a</sup> .....	4264	127252	30	57	21	1518
Lidzki .....	5605	186060	33	45	24	1936
Nowogródzki .....	2445	95907	39	62	13	696
Oszmiański .....	6885	189390	28	41	23	4435
Święciański .....	4826	139692	29	40	23	2695
Trocki .....	4160	92831	22	42	14	1231
Wilejski .....	6185	213424	35	41	28	2161
Wileński <sup>b</sup> .....	5727	313172	55	72	20	3799
<b>Brzeski</b> .....	<b>57758</b>	<b>1121978</b>	<b>19</b>	<b>33</b>	<b>189</b>	<b>5547</b>
Baranowicki .....	4495	178212	40	50	22	831
Brzeski <sup>c</sup> .....	4881	81497	17	54	24	491
Kobryński .....	5257	107814	21	43	28	650
Mozyrski .....	16155	240142	15	16	26	941
Piński .....	11867	222018	19	26	25	551
Prużański .....	4164	67343	16	41	21	522
Słonimski .....	7126	127737	18	40	24	767
Wołkowyski .....	3813	97215	25	50	19	794
<b>Miński</b> .....	<b>35947</b>	<b>1091138</b>	<b>30</b>	<b>38</b>	<b>85</b>	<b>8784</b>
Bobrujski .....	10500	188976	18	29	12	1030
Borysowski .....	6700	147673	22	31	17	1166
Ihumeński .....	8010	221758	28	32	21	1613
Miński <sup>d</sup> .....	5213	304004	58	67	16	3787
Stucki .....	5524	228727	41	43	19	1188

a–d W tym miasto: a – Grodno, b – Wilno, c – Brześć Litewski, d – Mińsk.

U w a g a. Nazwy okręgów zostały pogrubione.

Ź r ó d ł o: Romer (1920, s. 25).

**TABL. 2. LUDNOŚĆ WEDŁUG NARODOWOŚCI NA ZIEMIACH WSCHODNICH W GRUDNIU 1919 R.**

Okręgi, powiaty i wybrane miasta	Ogółem	Polacy	Białorusini	Litwini	Żydzi	Tutejsi	Inni
<b>O g ó ł e m</b> .....	<b>3846620</b>	<b>1218796</b>	<b>1651911</b>	<b>114595</b>	<b>392580</b>	<b>191981</b>	<b>276757</b>
<b>Wileński</b> .....	<b>1633504</b>	<b>874756</b>	<b>350391</b>	<b>112262</b>	<b>132603</b>	<b>37089</b>	<b>126403</b>
Brasławski .....	82513	43335	12622	12367	3254	–	10935
Dziśnieński .....	193263	74612	46381	1885	5280	–	65105
Grodzieński <sup>a</sup> .....	99087	53549	33160	1546	10534	–	298
Lidzki .....	186060	141479	27804	4203	6329	–	6245
Nowogródzki .....	95907	13833	72398	1	6974	–	2701
Oszmiański .....	189390	129165	11195	54	10639	33300	5037
Święciański .....	139692	61854	18727	38529	6524	2888	11170
Trocki .....	92825	49947	242	36748	3805	899	1184
Wilejski .....	213424	64549	123332	122	9453	2	15966
Wileński <sup>b</sup> .....	184218	160762	626	13864	6665	–	2301
Grodno <sup>c</sup> .....	28165	9255	2123	23	16587	–	177
Wilno <sup>c</sup> .....	128954	12416	1781	2920	46559	–	5278
<b>Brzeski</b> .....	<b>1121978</b>	<b>184334</b>	<b>598102</b>	<b>401</b>	<b>136259</b>	<b>116965</b>	<b>85917</b>
Baranowicki .....	178212	28568	110737	62	17972	4214	16559
Brzeski <sup>d</sup> .....	67492	14392	30543	–	7499	11957	3101
Kobryński .....	107814	16401	4554	83	9779	75557	1440
Mozyrski .....	240142	9599	188117	92	22841	3918	15575
Piński .....	222018	9675	151126	16	30160	6644	24397
Prużański .....	67343	7853	44841	48	11610	–	2991
Słonimski .....	127737	39416	49651	8	15094	3818	19750
Wołkowyski .....	97215	56173	17373	85	11269	10857	1458
Brześć Litewski <sup>e</sup> .....	14005	2257	1160	7	10035	–	546
<b>Miński</b> .....	<b>1091138</b>	<b>159706</b>	<b>703418</b>	<b>1932</b>	<b>123718</b>	<b>37927</b>	<b>64437</b>
Bobrujski .....	188976	9194	140600	572	27009	4752	6849
Borysowski .....	147673	32108	100075	219	9307	290	5674
Ihumeński .....	221758	26957	145625	109	15581	6493	26693
Miński <sup>e</sup> .....	201612	58237	98643	395	6661	26292	11344
Słucki .....	228727	14407	193066	2	18949	–	303
Mińsk <sup>c</sup> .....	102392	18713	23409	635	45961	100	13574

a Bez Grodna. b Bez Wilna. c Miasto. d Bez Brześcia Litewskiego. e Bez Mińska.

U w a g a. Jak przy tabl. 1. W publikacji źródłowej niektóre wartości nie sumują się do wartości ogółem.

Ź r ó d ł o: Romer (1920, s. 31).

**TABL. 3. WYBRANE NARODOWOŚCI WEDŁUG OKRĘGÓW W GRUDNIU 1919 R.**

Narodowości	Ogółem	Wileński	Brzeski	Miński
Rosjanie .....	220235	102022	57830	60383
Rusini .....	4129	–	4129	–
Tatarzy .....	3741	1316	381	2044
Inni .....	6890	3544	1927	1419
Nie podano .....	41762	19521	21650	591

Ź r ó d ł o: Romer (1920, s. 31).



## INNE CECHY CHARAKTERYSTYCZNE ZIEM WSCHODNICH WEDŁUG SPISU W 1919 R.

Rozmieszczenie ludności ziem wschodnich było bardzo niejednorodne. O ile na Litwie Środkowej, obejmującej powiaty: wileński, święciański, trocki i oszmiański, mieszkało 488968 osób na powierzchni 13490 km<sup>2</sup>, o tyle w pasie neutralnym między Litwą Środkową a Litwą Kowieńską<sup>5</sup> w powiatach wileńskim, trockim i święciańskim, obejmującym powierzchnię 1231 km<sup>2</sup>, mieszkało jedynie 21061 osób (GUS, 1921a, 1921b). Wynika z tego, że na Litwie Środkowej gęstość zaludnienia wynosiła ponad 36 osób na km<sup>2</sup>, podczas gdy w pasie neutralnym między Litwą Środkową a Litwą Kowieńską<sup>5</sup> gęstość zaludnienia wynosiła ok. 17 osób na km<sup>2</sup>, a więc była o ponad połowę niższa.

**TABL. 4. LUDNOŚĆ PASA NEUTRALNEGO POMIĘDZY LITWĄ ŚRODKOWĄ  
A LITWĄ KOWIEŃSKĄ W GRUDNIU 1919 R.**

Powiaty i gminy	Liczba ludności	Powiaty i gminy	Liczba ludności
<b>Pas neutralny ogółem</b> .....	<b>21061</b>	Wileński .....	15283
Trocki .....	5778	w tym:	
w tym:		Giedroycie .....	2902
Hanuszyski .....	439	Janiszki .....	1767
Jewle .....	1433	Mejszagola .....	200
Olkienniki .....	2041	Muśniki .....	2153
Rudziszki .....	726	Podbrzezie .....	82
Sumieliszki .....	226	Szyrwinty .....	5074
Troki .....	913	Święciański .....	3105

U w a g a. Nazwy gmin są podane w formie rzeczownikowej.

Ź r ó d ł o: GUS (1921b).

**TABL. 5. LUDNOŚĆ LITWY ŚRODKOWEJ WEDŁUG SPISU W GRUDNIU 1919 R.**

Powiaty, gminy i wybrane miasta	Liczba ludności	Powiaty, gminy i wybrane miasta	Liczba ludności
<b>Litwa Środkowa ogółem</b> .....	<b>488968</b>	Rzesza .....	11986
Wileński .....	120362	Sołeczniaki .....	6362
Wilno <sup>a</sup> .....	128954	Szumsk .....	7322
Bystrzyca .....	6355	Worniany .....	19408
Ilino .....	12797	Święciański .....	130182
Janiszki .....	4600	Święciany <sup>a</sup> .....	3900
Mejszagola .....	9859	Nowo-Święciany <sup>a</sup> .....	3277
Mickuny .....	8074	Aleksandrowo .....	6581
Niemenczyn .....	9906	Dubatówka .....	3156
Podbrzezie .....	7550	Dangieliszki .....	10268
Rudomino .....	9009	Hoduciszki .....	8397
Rukojmie .....	7134	Jasiew .....	2181

a – Miasto.

U w a g a. Jak przy tabl. 4.

<sup>5</sup> Pas neutralny między Polską a Litwą o szerokości 10–12 km, wytyczony w lipcu 1919 r. przez marszałka Ferdynanda Focha i zatwierdzony w grudniu 1919 r. przez Radę Najwyższą Ententy, nie należał wówczas ani do Polski, ani do Litwy. Zlikwidowano go dopiero w lutym 1923 r., kiedy ustalono ostateczne granice obu państw (Łossowski, 2002, s. 28).

**TABL. 5. LUDNOŚĆ LITWY ŚRODKOWEJ WEDŁUG SPISU W GRUDNIU 1919 R. (dok.)**

Powiaty, gminy i wybrane miasta	Liczba ludności	Powiaty, gminy i wybrane miasta	Liczba ludności
Kiemieliszki .....	10214	Jewle .....	420
Komaje .....	5465	Landwarów .....	9256
Kobylniki .....	4952	Olkieniki .....	7052
Łyngmiany .....	5100	Rudziszki .....	5878
Łyntupy .....	7356	Troki .....	6144
Melegjany .....	5880	Oszmiański .....	78834
Michałowo .....	8779	Oszmiana <sup>a</sup> .....	3442
Świr .....	8708	Bienica .....	7957
Święciany .....	7888	Dzieweniszki .....	11185
Szemietowszczyzna .....	2986	Grauzyszki .....	7525
Twerecz .....	4347	Holszany .....	10001
Wojstom .....	7066	Krewo .....	4102
Wiszniew .....	4603	Kucewicze .....	6481
Zanarocz .....	1553	Polany .....	14130
Zabłociszki .....	7525	Smorgonie .....	3860
Trocki .....	30636	Soły .....	10151
Troki <sup>a</sup> .....	1886		

a – Miasto.

U w a g a. Jak przy tabl. 4.

Ź r ó d ł o: GUS (1921a).

Mała liczba ludności miast powiatowych na ziemiach wschodnich w grudniu 1919 r. wynikała ze strat ludności i migracji na skutek prowadzonych działań wojennych (Holzer i Molenda, 1963; Romer, 1920). Przedstawiała się ona następująco w poszczególnych okręgach:

- wileński: Braśław – 1235, Dzisna – 4445, Grodno – 28165, Lida – 11365, Nowogródek – 5096, Oszmiana – 3442, Święciany – 3900, Troki – 1886, Wilejka – 5893, Wilno – 128954;
- brzeski: Baranowicze – 10373, Brześć Litewski – 14005, Kobryń – 8835, Mozyrz – 6796, Pińsk – 21436, Prużana – 6173, Słonim – 9719, Wołkowysk – 8472;
- miński: Bobrujsk – 29704, Borysów – 5452, Ihumeń – 4669, Mińsk – 102392, Słuck – 14162.

Członek amerykańskiej technicznej Misji Doradczej prof. E. Durand z Uniwersytetu Minnesota z siedzibą w Minneapolis, będący ekspertem w sprawach spisowych, bardzo wysoko ocenił jakość danych spisu z grudnia 1919 r. przeprowadzonego na ziemiach wschodnich przez CBSL w Wilnie. Między innymi na podstawie tej rekomendacji część wyników tego spisu włączono do wyników spisu ogólnopolskiego z 1921 r.

Niewielkich, choć ważnych, korekt spisu z grudnia 1919 r. dokonali Wakar i Szturm de Sztrem w analitycznym artykule opublikowanym w czwartym numerze „Ekonomisty” z 1920 r. Zwrócili uwagę, że spis został przeprowadzony głównie przez urzędników w ciągu kilku tygodni w sposób „dosyć” formalny, co ich zdaniem mogło spowodować stronniczość przy opisie narodowości mieszkańców ziem wschodnich, którzy nie mieli poczucia przynależności narodowej.

Notowani w spisie tutejsi to według Szturm de Sztrema i Wakara w większości Białorusini i Rusini (Szturm de Sztrem i Wakar, 1920).

Współczesne badania dowodzą, że mieszkańcy Polesia (Poleszucy) dopiero po II wojnie światowej zaczęli postrzegać swoją grupę w kategoriach tożsamości narodowej. W spisach z okresu II Rzeczypospolitej określali się w większości jako tutejsi, a swój język nazywali prostym (Ciesielski, 2004; Tomaszewski, 1963, 1985). Współczesna Białoruś uznaje ich za odrębną grupę etniczną.

Niezwykłe zróżnicowany skład narodowościowy ziem wschodnich, w tym znaczny odsetek osób deklarujących się jako tutejsi, był jednym z ważniejszych problemów, z którym musiało zmierzyć się państwo polskie na terytorium historycznej Litwy.

## PODSUMOWANIE

Spis ludności przeprowadzony w grudniu 1919 r. przez Komisariat Generalny Zarządu Cywilnego Ziem Wschodnich (ZCZW) wykazał, że na terytorium Litwy historycznej mieszkało wówczas 3846620 osób. Odnotowano jedynie osiem miast powyżej 10 tys. mieszkańców i bardzo duże zróżnicowanie wyznaniowe. Największą grupą narodowościową zamieszkującą ziemie wschodnie byli Białorusini, następnie: Polacy, Żydzi, Rosjanie, tutejsi, Litwini, Rusini i Tatarzy. Poszczególne okręgi również charakteryzowały się znacznym zróżnicowaniem narodowościowym. O ile w okręgu wileńskim przeważali Polacy, a drugą co do wielkości grupą narodowościową byli Białorusini, o tyle w okręgach brzeskim i mińskim większość stanowili Białorusini, a Polacy – drugą co do wielkości grupę narodowościową. Jedynie w miastach okręgu wileńskiego dominowali Polacy, natomiast w miastach okręgu brzeskiego i mińskiego – zdecydowanie Białorusini.

Spis miał duże znaczenie dla odbudowy polskiej państwowości na tych terenach. Na podstawie danych dostarczonych do Ministerstwa Spraw Wewnętrznych przez Wydział Statystyczny byłego ZCZW oraz byłego Zarządu Cywilnego Ziem Wołynia i Frontu Podolskiego możliwe stało się m.in. opracowanie tymczasowego podziału administracyjnego ziem wcielonych do Rzeczypospolitej Polskiej po podpisaniu w 1921 r. w Rydze pokoju z Rosją Radziecką. Znaczna część wyników spisu z 1919 r. na ziemiach wschodnich została włączona do wyników pierwszego powszechnego spisu ludności przeprowadzonego przez GUS we wrześniu 1921 r. Poznanie stanu ludności ziem wschodnich uzyskane po spisie 1919 r. pozwoliło także na zaplanowanie długofalowego programu osadnictwa wojskowego na tych terenach, mającego na celu zmniejszenie dystansu cywilizacyjnego ziem wschodnich w stosunku do centralnej Polski i umocnienie roli państwa polskiego na obszarze Litwy historycznej (Cichoracki, 2014; Śleszyński, 2004; Śleszyński i Szarejko, 2018).

## BIBLIOGRAFIA

- CBSL. (1920a). *Grafik składu wyznaniowego narodowości [i] składu narodowościowego wyznań*. Wilno: Komisariat Generalny ZCZW [Zarządu Cywilnego Ziem Wschodnich], Centralne Biuro Spisu Ludności [rękopis].
- CBSL. (1920b). *Grafik wyznaniowy [i] narodowościowy w procentach: okręgu: wileńskiego, brzeskiego, mińskiego łącznie z miastami, powiatów bez miast powiatowych, miast powiatowych; terenu objętego ZCZW* [Zarządem Cywilnym Ziem Wschodnich]. Wilno: Centralne Biuro Spisu Ludności [rękopis].
- Cichoracki, P. (2014). Polonisation projects for Polesia and their delivery in 1921–1939. *Acta Poloniae Historica*, (109), 61–79.
- Ciesielski, S. (2004). Kresy Wschodnie II Rzeczypospolitej i problemy identyfikacji narodowej. W: S. Ciesielski (red.), *Przemiany narodowościowe na Kresach Wschodnich II Rzeczypospolitej 1931–1948* (s. 9–51). Toruń: Wydawnictwo Adam Marszałek.
- Czekanowski, J. (1919). *Note sur la statistique démographique des confins orientaux de la Pologne*. Referat wygłoszony na Conférence de la paix, Paris.
- Dmitruk, S. (2017). Ludność ewakuowana z zachodnich powiatów guberni grodzieńskiej na Syberii w 1916 r. na przykładzie guberni tomskiej. *Przeszłość Demograficzna Polski*, 39, 283–322. DOI: 10.18276/pdp.2017.39-11.
- Dominiczak, H. (1992). *Granica wschodnia Rzeczypospolitej Polskiej w latach 1919–1939*. Warszawa: Wydawnictwo Naukowe PWN.
- Eberhardt, P. (1999). Kresy Wschodnie – granice, terytorium, ludność polska. W: K. Jasiewicz (red.), *Europa nieprowincjonalna. Przemiany na ziemiach wschodnich dawnej Rzeczypospolitej (Białoruś, Litwa, Łotwa, Ukraina, wschodnie pogranicze III Rzeczypospolitej Polskiej w latach 1772–1999)* (s. 29–38). Warszawa: Rytm.
- Gierałowska-Kałużur, J. (2002). Zarząd Cywilny Ziem Wschodnich (19 lutego 1919 r.–9 września 1920 r.). *Dzieje Najnowsze*, 34(3), 141–148.
- Gierałowska-Kałużur, J. (2003). *Zarząd Cywilny Ziem Wschodnich (19 lutego 1919 r.–9 września 1920 r.)*. Warszawa: Wydawnictwo Neriton, Instytut Historii PAN.
- GUS. (1921a). Ludność Litwy Środkowej wg spisu ludności z r. 1919. *Miesięcznik Statystyczny*, 4–6, 104.
- GUS. (1921b). Ludność pasa neutralnego pomiędzy Litwą Środkową a Litwą Kowieńską wg spisu ludności r. 1919. *Miesięcznik Statystyczny*, 4–6, 105.
- Holzer, J., Molenda, J. (1963). *Polska w Pierwszej Wojnie Światowej*. Warszawa: Wiedza Powszechna.
- Kirwiel, E. (2006). Cechy charakterystyczne Kresów Północno-Wschodnich Rzeczypospolitej Polskiej okresu międzywojennego. *Annales Universitatis Mariae Curie-Skłodowska*, 13, 93–105.
- Koter, M. (1997). Kresy państwowe – geneza i właściwości w świetle doświadczeń geografii politycznej. W: K. Handke (red.), *Kresy – pojęcie i rzeczywistość. Zbiór studiów* (s. 44–51). Warszawa: Sławistyczny Ośrodek Wydawniczy.
- Kozyra, W. (2005). Polityka administracyjna władz polskich na Ziemiach Wschodnich Rzeczypospolitej Polskiej w latach 1918–1926. *Annales Universitatis Mariae Curie-Skłodowska, Sectio F, Historia*, 60, 419–440.
- Łossowski, P. (2002). Ostatni akt kształtowania się granic Drugiej Rzeczypospolitej – podział pasa neutralnego pomiędzy Polską a Litwą w lutym 1923 r. *Studia z Dziejów Rosji i Europy Środkowo-Wschodniej*, 37, 27–39.
- Majchrowski, J. M. (1994). *Kto był kim w Drugiej Rzeczypospolitej*. Warszawa: BGW.

- Maliszewski, E. (1920). Polski spis ludności na Ziemiach Wschodnich z grudnia 1919 r. *Wschód Polski*, 10/11, 14–21.
- Mielcarek, A. J. (2008). *Podziały terytorialno-administracyjne II Rzeczypospolitej w zakresie administracji zespolonej*. Warszawa: Wydawnictwo Neriton.
- Mielcarek, A. J. (2011). Węzłowe zagadnienia ustrojowe Zarządu Cywilnego Ziem Wschodnich (1919–1920) w świetle aktów normatywnych. *Studia z Dziejów Państwa i Prawa Polskiego*, 14, 241–251.
- Mironowicz, E. (2007). *Białoruś*. Warszawa: Wydawnictwo Trio.
- Nowak, A. (1993). *Wojna polsko-sowiecka 1919–1921 roku w świetle najnowszych publikacji*. *Kwartalnik Historyczny*, 100(3), 79–107.
- Olszewska, R., Dłuska, T. (1981). *Tematyka i organizacja spisów powszechnych w Polsce*. Warszawa: Główny Urząd Statystyczny.
- Praesent, H. (1919). Kriegsmäßige Volkszählungen im Generalgouvernement Warschau und die Bevölkerungszahl in Kongreß-Polen. *Jahrbücher für Nationalökonomie und Statistik*, 113(1), 52–64.
- Romer, E. (1920). *Spis ludności na terenach administrowanych przez Zarząd Cywilny Ziem Wschodnich (grudzień 1919)*. Lwów–Warszawa: Książnica Polska Towarzystwa Nauczycieli Szkół Wyższych.
- Romer, E., Szumański, T. (1920). *Mapa etnograficzna Ziem Wschodnich (Polska, Litwa, Białoruś) podług spisu oficjalnego z grudnia 1919*. Lwów: Lit. Akc. Tow.
- Rzepkowski, A. (2005). Spisy ludności na ziemiach polskich w latach 1789–1939. *Przegląd Nauk Historycznych*, 4(2), 101–125.
- Siemakowicz, M. (1998). Spisy ludności a zagadnienie narodowościowe z uwzględnieniem spraw szkolnictwa dla mniejszości białoruskiej w II Rzeczypospolitej. *Białoruskie Zeszyty Historyczne*, 10, 82–106.
- Strzelecki, Z., Toczyński, T. (2002). *Spisy ludności Rzeczypospolitej Polskiej 1921–2002*. Warszawa: Polskie Towarzystwo Demograficzne, Główny Urząd Statystyczny.
- Szturm de Sztrem, T., Wakar, W. (1920). Spis ludności na Litwie. *Ekonomista*, 4, 33–36.
- Szturm de Sztrem, T., Wakar, W. (1921). *Stosunki narodowościowe na terenach b. Zarządu Cywilnego Ziem Wschodnich: według spisu z 1919 r.* Warszawa: Wydawnictwo Instytutu Gospodarstwa Społecznego.
- Śleszyński, W. (red.). (2009). *Polesie w polityce rządów II Rzeczypospolitej*. Białystok–Kraków: Instytut Historii Uniwersytetu w Białymstoku, Wydawnictwo Avalon.
- Śleszyński, W., Szarejko, M. (2018). *Niepodległa: ziemie wschodnie II Rzeczypospolitej*. Białystok: Muzeum Pamięci Sybiru.
- Thugutt, S. (1915). *Polska i Polacy. Ilość i rozsiadlenie ludności polskiej*. Warszawa: Polskie Towarzystwo Krajoznawcze.
- Tomaszewski, J. (1963). *Z dziejów Polesia 1921–1939. Zarys stosunków społeczno-ekonomicznych*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Tomaszewski, J. (1985). *Ojczyzna nie tylko Polaków. Mniejszości narodowe w Polsce w latach 1918–1939*. Warszawa: Młodzieżowa Agencja Wydawnicza.

## Mirosław Szreder *Zrozumieć świat liczb* (Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk 2019)

Na rynku księgarskim niewiele jest książek, które w przystępny sposób przedstawiają statystykę nie jako zbiory liczb ujętych w tablice i diagramy, lecz jako pasjonującą naukę o otaczającym nas świecie. Współczesną statystykę postrzega się często przez pryzmat liczb, wskaźników i indeksów, wskutek czego nie zawsze dostrzega się to, co w niej najistotniejsze. Zdaniem Mirosława Szredera głównym zadaniem statystyki jest dostarczenie innym naukom, a także szerokiemu gronu odbiorców narzędzi do wydobywania wartościowej wiedzy z coraz bogatszych zbiorów liczb. W przedmowie do książki *Zrozumieć świat liczb* autor zapowiada podzielenie się z czytelnikami refleksją na temat możliwości zdobywania nowej wiedzy, jakie oferuje dzisiaj statystyka. Chce przedstawić zarówno możliwości statystyki i rachunku prawdopodobieństwa, jak i zagrożenia związane z rosnącymi w sposób niekontrolowany zasobami danych.

Na książkę składa się łącznie 27 tekstów naukowych i popularnonaukowych, z których większość ukazała się w ostatnich latach w poświęconych nauce działach „Polityki” i „Tygodnika Powszechnego” oraz na łamach „Rzeczpospolitej”. Wiele z tych artykułów przybliżyło mało znane formy myślenia statystycznego i zastosowań metod statystycznych. Zaletą publikacji jest jasny i zrozumiały język, pozwalający dotrzeć do każdego czytelnika, który chciałby poznać tajemniczy dla wielu osób warsztat badawczy statystyków. Prawie wszystkie teksty dotyczą konkretnych i aktualnych zagadnień społecznych i gospodarczych, dla których zrozumienia lub rozwiązania pomocna może się okazać statystyka.

Książka podzielona jest na trzy części: *Po co nam pomiar liczbowy i statystyka?*, *Liczb, prawdopodobieństwo i ryzyko spotkania czarnych łabędzi* oraz *Sondaże, emocje i wybory*. Każda z nich liczy od 8 do 11 artykułów, których tematykę dobrze oddają tytuły odpowiednich części. Jedyne opracowanie z części I zatytułowane *Gwałt na języku w nauce*, które traktuje o sprawach języka polskiego i żargonu naukowego, nie jest bezpośrednio związane z pomiarem liczbowym ani statystyką. Odnosi się ono do wszystkich dyscyplin naukowych i niebezpiecznej tendencji tworzenia przez część naukowców hermetycznego, niezrozumiałego dla innych ludzi języka.

Część I składa się głównie z artykułów, w których autor podejmuje i dyskutuje podstawowe kwestie związane z pomiarem liczbowym, będącym niejako warunkiem wstępnym zastosowań statystyki. Ta część książki rozpoczyna się od ciekawego tekstu, który – być może ku zaskoczeniu czytelnika – nie jest wcale apologią statystyki, lecz zwraca uwagę zarówno na zalety, jak i zagrożenia oraz pułapki liczbowego pomiaru. W drugim artykule *Prawda liczb i statystyk* autor sięga jeszcze głębiej i prezentuje tę gałąź nauki w sposób umożliwiający wyro-

bień własnego poglądu na znaczenie badań statystycznych we współczesnym świecie. Trzecie opracowanie zostało poświęcone fenomenowi losowania i rosnącej popularności zastosowań losowego wyboru w różnych dziedzinach życia, m.in. w nauce (losowanie recenzentów) oraz wymiarze sprawiedliwości (losowanie sędziów). Warto przypomnieć, że losowanie jest nierozzerwalnie związane z reprezentacyjnymi badaniami statystycznymi, w których wnioskowanie opiera się na próbie losowej. Z kolei *Lekcje zdrowia* podejmują mało znany temat współzależności między wykształceniem a długością życia oraz zadowoleniem. Piąty artykuł w tej części to jasny i przystępny wykład na temat wad i zalet średniej arytmetycznej i mediany jako mierników przeciętnych płac w gospodarce. Czytelnik dowie się z niego m.in., dlaczego nie jest łatwo posługiwać się medianą zamiast krytykowaną przez wiele osób średnią arytmetyczną płac. Tekst *Umysł w epoce postprawdy* nawiązuje do skutków błędów poznawczych, które wszyscy popełniamy, a które szczegółowo zbadał i opisał Daniel Kahneman (laureat Nagrody Nobla w dziedzinie ekonomii). Dla Szredera inspiracją do napisania tego tekstu były jednak nie czyjeś nieuświadomione błędy, lecz pojawienie się w sferze publicznej tzw. faktów alternatywnych i fake newsów. Artykuł *Wielki jeden procent* pokazuje, jak wiele można stracić w opisie różnych niehomogenicznych zbiorowości, gdy pominię się małą, np. jednoprocentową, frakcję jednostek takich zbiorowości. Jest to – moim zdaniem – ostrzeżenie, aby wiara w duże zbiory danych (np. big data) nie przesłaniała istotnego wpływu, jaki mogą mieć pominięte jeden lub kilka procent obserwacji. Klamrę spinającą pierwszą część książki stanowi tekst o szkodliwej dla nauki i jej postrzegania w społeczeństwie tendencji do tworzenia nowych, niezrozumiałych kategorii i pojęć lub niewłaściwego ich używania.

Część II zawiera opracowania traktujące o nieco bardziej złożonych kwestiach pomiaru liczbowego, w tym o prawdopodobieństwie i jego interpretacjach, ryzyku, a także najbardziej aktualnych pod tym względem trendach, do których z pewnością należy zaliczyć zjawisko big data. Dwa pierwsze artykuły to próba wyjaśnienia nietrywialnych spraw interpretowania i rozumienia prawdopodobieństwa jako matematycznej miary niepewności. Zawierają ciekawe przykłady ilustrujące popełniane przez wiele osób błędy w pojmowaniu samego prawdopodobieństwa, jak również operacji przeprowadzanych na prawdopodobieństwach. Trzy kolejne teksty przedstawiają wyzwania, przed którymi stoi nie tylko nauka, lecz także życie codzienne społeczeństwa wobec rosnących możliwości gromadzenia i analizowania wielkich zbiorów danych. W artykule *Czy lepiej nie wiedzieć? (O algorytmach Big Data)* autor zawarł krótkie omówienie książki *Broń matematycznej zagłady*, a jednocześnie polemikę z jej autorką Cathy O'Neil *Jak algorytmy zwiększają nierówność i zagrażają demokracji*. Niekoniecznie pozytywny wpływ zwiększającej się liczby danych, które może zdobyć badacz, na możliwość zastosowań wnioskowania statystycznego został przedstawiony w opracowaniu *Gdy więcej oznacza mniej*. Opracowanie to poprzedza artykuł o równie

oryginalnym tytule *Czarne łabędzie*. Myliłby się jednak ten, kto spodziewałby się omówienia lub nawiązania do głośnej książki Nassima Nicholas Taleba *Czarny łabędź*. Szreder odwołuje się do innych źródeł literackich, by przedstawić swoje spojrzenie na zdarzenia o bardzo małym prawdopodobieństwie wystąpienia, ale bardzo poważnych konsekwencjach – właśnie takie zdarzenia przyjęło się wśród ekonomistów i strategów określać mianem *black swans*, czyli „czarnych łabędzi”. Autor podejmuje tę problematykę w kontekście ryzyka i sposobów jego minimalizowania. Oczywiście jest tu także miejsce na przedyskutowanie kwestii szacowania prawdopodobieństwa rzadkich (niepowtarzalnych) zdarzeń, co łączy to opracowanie z wcześniej wspomnianymi artykułami o prawdopodobieństwie. Uważam, że *Czarne łabędzie* mogą stanowić literaturę uzupełniającą z zakresu analizy ryzyka lub zarządzania ryzykiem dla studentów różnych kierunków studiów, nie tylko ekonomicznych.

Część III książki została w całości poświęcona problematyce sondaży, czyli badań statystycznych, w których losowa próba respondentów stanowi podstawę do wnioskowania o opiniach lub postawach większych zbiorowości. Autora interesują przede wszystkim sondaże ogólnopolskie, czyli zbiorowości wszystkich wyborców albo wszystkich dorosłych Polaków. Szreder podkreśla rolę statystyki, mimo że dla wielu czytelników pierwszą dziedziną nauki, z jaką wiązałyby sondaże, byłaby pewnie socjologia. W mediach wyniki sondaży komentują przecież przede wszystkim socjologowie. Niemniej od kilku lat Szreder, czyli właśnie statystyk, jest jednym ze stałych komentatorów sondaży w „Rzeczpospolitej”, przy czym jego opinie odnoszą się nie do zagadnień politycznych czy społecznych, lecz jedynie do podstaw metodologicznych (statystycznych) tego rodzaju badań. Wszystkie artykuły w tej części książki, a jest ich 11, dotyczą metody projektowania i realizowania badań sondażowych. W większości z nich pretekstem do wytłumaczenia czytelnikom zawichości „kuchni statystycznej” w tego rodzaju badaniach były porażki w prognozowaniu wyników wyborczych w Polsce i na świecie. Dużo jest w tych opracowaniach krytyki pod adresem realizatorów badań, czyli pracowni sondażowych, a także mediów, które zdaniem autora są głównym klientem ośrodków badawczych. Media stanowią przedmiot tekstu *Stadna pułapka*, w którym Szreder doszukuje się możliwości podążania przez pracowni sondażowe za aktualnymi trendami na rynku (zjawisko przyłączania się do stada), co prawdopodobnie miało miejsce m.in. w Wielkiej Brytanii podczas kampanii wyborczej w 2015 r. Chociaż gdy przyjrzeć się nietrafionym wynikom sondażowym, okaże się, że przyczyniają się do nich także respondenci. *Powiedz prawdę, respondencie* traktuje właśnie o dylematach respondentów takich badań. We wszystkich tekstach składających się na część III książki widoczny jest duży zapał autora do wyjaśniania czytelnikom założeń metodycznych, jakie legły u podstaw rozwinięcia w latach 30. ubiegłego wieku badań sondażowych, a które zdają się być współcześnie u szczytu popularności. Jednak, jak stwierdza autor, ta popularność doprowadziła m.in. do tego, że codzienne sondaże



mają coraz mniej wspólnego z prawdziwą nauką. Wykonuje się je mało starannie i pod dużą presją czasu. Dlatego Szreder proponuje konkretne działania, które mogłyby zatrzymać pogarszanie się reputacji sondaży (*Naprawianie sondaży*) i ostatecznie ich broni (*Sondaże służą czy szkodzą*). Podpisuję się pod tytułowym apelem autora z ostatniego tekstu: *Nie zakazujmy sondaży!*. Mimo że bywają one zawodne, stanowią wciąż jeden z najpopularniejszych rodzajów badań statystycznych próbkowych, z którymi na co dzień mamy do czynienia. Warto je doskonalić zamiast z nich rezygnować lub zabraniać publikowania ich wyników.

Omawiana książka ma oryginalną szatę graficzną, wzbogaconą reprodukcjami obrazów Romana Opałki (1931–2011) z cyklu tzw. obrazów liczonych oraz krótkim wspomnieniem artysty autorstwa Wojciecha Szafrąńskiego.

Uważam, że książka Szredera *Zrozumieć świat liczb* może stać się cenną pozycją w literaturze uzupełniającej do przedmiotów z zakresu m.in. statystyki, badań socjologicznych i badań medioznawczych. Dla osób spoza środowiska akademickiego będzie z pewnością stanowiła zbiór interesujących opracowań na temat rozumienia zjawisk opisywanych za pomocą liczb.

**Zenon Wiśniewski** (Uniwersytet Mikołaja Kopernika w Toruniu)

## Wydawnictwa GUS. Wrzesień 2019



We wrześniowej ofercie wydawniczej warto zwrócić uwagę na cykliczną (ukazującą się co trzy lata) publikację poświęconą działalności uniwersytetów trzeciego wieku. Tegoroczne opracowanie pt. **Uniwersytety Trzeciego Wieku w roku akademickim 2017/2018** przedstawia wyniki drugiej edycji ogólnopolskiego badania pn. Sprawozdanie z działalności Uniwersytetów Trzeciego Wieku, przeprowadzonego w 2018 r. Działalność tych instytucji jest ważnym elementem polityki senioralnej, wpisującym się w ideę społeczeństwa opartego na wiedzy i w koncepcję uczenia się przez całe życie.

Publikacja zawiera bogaty zestaw informacji o uniwersytetach trzeciego wieku, z uwzględnieniem m.in. ich formy organizacyjnoprawnej, roku rozpoczęcia działalności, województwa, w którym dana placówka funkcjonuje, oraz głównej idei jej przyświecającej. Scharakteryzowano także słuchaczy oraz osoby, które w ramach tych instytucji uczestniczą w działaniach edukacyjnych, aktywizacyjnych i integracyjnych (m.in. według poziomu wykształcenia i statusu na rynku pracy). Podano również źródła finansowania oraz informacje o współpracy z podmiotami zewnętrznymi i o czynnikach utrudniających prowadzenie działalności.

Opracowanie, wydane w wersji polsko-angielskiej, jest dostępne na stronie internetowej GUS.

We wrześniu br. ukazały się ponadto:

- „Biuletyn statystyczny” nr 8/2019;
- *Bezrobocie rejestrowane. I–II kwartał 2019 roku;*
- *Budżety gospodarstw domowych w 2018 r.;*
- *Ceny robót budowlano-montażowych i obiektów budowlanych (lipiec 2019 r.);*
- *Efekty działalności budowlanej w 2018 r.;*
- *Koniunktura w przetwórstwie przemysłowym, budownictwie, handlu i usługach 2000–2019 (wrzesień 2019);*
- *Kultura w 2018 r.;*
- *Nakłady i wyniki przemysłu – I–II kwartał 2019 r.;*
- *Poczta i telekomunikacja – wyniki działalności w 2018 r.;*
- *Pomoc społeczna i opieka nad dzieckiem i rodziną w 2018 r.;*
- *Produkcja ważniejszych wyrobów przemysłowych w sierpniu 2019 r.;*
- *Sytuacja społeczno-gospodarcza kraju w sierpniu 2019 r.;*
- *Sytuacja społeczno-gospodarcza województw Nr 2/2019;*

- *Sytuacja makroekonomiczna w Polsce na tle procesów zachodzących w gospodarce światowej w 2018 r.;*
- *Transport – wyniki działalności w 2018 r.;*
- *Zatrudnienie i wynagrodzenia w gospodarce narodowej w pierwszym półroczu 2019 r.;*
- „Wiadomości Statystyczne” nr 9/2019.

Wersje elektroniczne wszystkich publikacji GUS są dostępne 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” otrzymała 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 przysyłać drogą elektroniczną na adres: [redakcja.ws@stat.gov.pl](mailto: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 utajeniu 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-

Redakcja 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](http://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 opisanie 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 opóźnić 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 jakiegokolwiek 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 przepras.

### **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 punktoremami 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. **P**, **N<sub>ij</sub>**); wektory – małe litery, kursywa, pogrubione (np. **w**, **x<sub>i</sub>**); pozostałe zmienne – małe lub duże litery, kursywa, bez pogrubienia (np. *w*, *x<sub>i</sub>*, *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 – wykrs.
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 obcojęzycznych.

- 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](http://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 znacząco 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.