Cena zł 12,00
(VAT 5%)

Indeks 381306
PL ISSN 0043-518X

WIADOMOŚCI STATYSTYCZNE

GŁÓWNY
URZĄD
STATYSTYCZNY

POLSKIE
TOWARZYSTWO
STATYSTYCZNE

MIESIĘCZNIK
ROK LX
WARSZAWA
STYCZEŃ 2015

1

REDAKCJA

al. Niepodległości 208, 00-925 Warszawa, gmach GUS, pok. 353, tel. 22 608-32-25
http://www.stat.gov.pl/pts/16_PLK_HTML.htm

Elżbieta Grabowska (e.grabowska@stat.gov.pl)

Wersja internetowa jest wersją pierwotną czasopisma.

ZAKŁAD WYDAWNICTW STATYSTYCZNYCH

al. Niepodległości 208, 00-925 Warszawa, tel. 22 608-31-45.
Informacje w sprawach nabywania czasopism tel. 22 608-32-10, 608-38-10.
Zbigniew Karpiński (redaktor techniczny), Ewa Krawczyńska (skład i łamanie),
Wydział Korekty pod kierunkiem Bożeny Gorczycy, mgr Andrzej Kajkowski (wykresy).

Indeks 381306

Prenumerata realizowana przez RUCH S.A.:

Zamówienia na prenumeratę w wersji papierowej i na e-wydania można składać bezpośrednio na stronie www.prenumerata.ruch.com.pl.
Ewentualne pytania prosimy kierować na adres e-mail: prenumerata@ruch.com.pl lub kontaktując się z Centrum Obsługi Klienta „RUCH" pod numerami: 22 693-70-00 lub 801 800 803 — czynne w dni robocze w godzinach 7^{00}—17^{00}.
Koszt połączenia wg taryfy operatora.

Zam. 532/2014 — nakład 420 egz.

W dniu 23 grudnia 2014 r. zmarł

prof. dr hab. Tadeusz Walczak

Wiceprezes Głównego Urzędu Statystycznego w latach 1972–1990,
Dyrektor Ośrodka Elektronicznego GUS w latach 1967–1972,
od 1992 r. Radca Prezesa GUS,
emerytowany Profesor Szkoły Głównej Handlowej w Warszawie.

Pan Profesor Tadeusz Walczak był jednym z najwybitniejszych polskich autorytetów naukowych w dziedzinie metodologii i projektowania badań statystycznych, ochrony danych w systemach informacyjnych oraz zastosowań informatyki w statystyce.

Przez blisko 60 lat pracy w GUS Profesor wniósł fundamentalny wkład w rozwój i popularyzację statystyki publicznej. Był aktywnym członkiem Naukowej Rady Statystycznej, Międzynarodowego Instytutu Statystycznego,

Komitetu Statystyki i Ekonometrii PAN, Rady Głównej PTS oraz redaktorem naczelnym miesięcznika „Wiadomości Statystyczne".

Profesor był autorem ponad 200 publikacji i artykułów z zakresu wykorzystania informatyki w gospodarce, metodologii i projektowania badań statystycznych, historii statystyki oraz historii GUS.

Ogromny dorobek naukowy oraz zawodowy stanowił podstawę uhonorowania Profesora wieloma wysokimi odznaczeniami państwowymi i resortowymi, m.in.: Krzyżem Komandorskim Orderu Odrodzenia Polski, Krzyżem Oficerskim Orderu Odrodzenia Polski, Krzyżem Kawalerskim Orderu Odrodzenia Polski, Złotym Krzyżem Zasługi.

Pan Profesor Tadeusz Walczak zapisał się w historii GUS jako wybitny statystyk służący swoją wiedzą i doświadczeniem. Będziemy wspominać Profesora jako Człowieka wielkiej prawości, niezmiennie życzliwego i pomocnego zarówno w relacjach prywatnych, jak i zawodowych.

Przepełnieni poczuciem ogromnej straty

Prezes, kierownictwo i pracownicy
Głównego Urzędu Statystycznego

NR 1 (644)
STYCZEŃ 2015

WIADOMOŚCI STATYSTYCZNE

CZASOPISMO GŁÓWNEGO URZĘDU STATYSTYCZNEGO
I POLSKIEGO TOWARZYSTWA STATYSTYCZNEGO

Pożegnanie Pana Profesora Tadeusza Walczaka

**Wielce Szanowny Panie Profesorze, Panie Prezesie,
Drogi Tadeuszu i Przyjacielu**

Z wielkim zaskoczeniem i niedowierzaniem przyjęliśmy wiadomość o śmierci Profesora Tadeusza Walczaka, która jest niezwykle przygnębiająca i przepełnia nas wielkim żalem. To bardzo smutny dzień dla naszej statystycznej rodziny, do której Profesor należał od tak dawna.

Wiadomość nieoczekiwana, gdyż praktycznie do ostatniej chwili Profesor był z nami, pracował intensywnie i wspólnie planowaliśmy dalszą współpracę nad kolejnymi ważnymi zadaniami. Przez całe życie zawodowe Profesor realizował zadania pionierskie, niezwykle ważne dla sprawnego funkcjonowania statystyki. Dotyczyły one wielu obszarów, ale przede wszystkim zastosowania informatyki w praktyce badań statystycznych. Jako dyrektor Ośrodka Elektronicznego GUS oraz wieloletni wiceprezes GUS — Profesor organizował, wdrażał i odpowiadał za zastosowanie i sprawne funkcjonowanie informatyki w statystyce. Można powiedzieć, że Profesor był przy narodzinach informatyki w GUS i uczestniczył w kolejnych fazach jej rozwoju, aż do współczesnego znaczenia systemów informatycznych.

Wiadomość o odejściu Profesora była dla nas niezwykle smutna i przygnębiająca, gdyż straciliśmy nie tylko znakomitego statystyka, naukowca, badacza, informatyka i popularyzatora statystyki w kraju i za granicą, ale przede wszystkim wspaniałego człowieka, kolegę i przyjaciela. Profesor był człowiekiem ogromnej życzliwości, niespotykanej skromności, wyróżniającym się niezwykłą kulturą osobistą, a jednocześnie otwartym na współpracę i gotowość pomocy innym. Doświadczaliśmy tego stale, a dla mnie postawa życiowa i zawodowa Profesora była nie tylko wzorem, ale również niezwykłym wsparciem. Wiedziałem, że zawsze mogę na Profesora liczyć, że w potrzebie nie odmówi mi swoje-

go wsparcia organizacyjnego, merytorycznego czy po prostu przyjacielskiej rady. To Jego zaangażowanie i obecność dodawała nam sił i pewności w działaniu. Był przecież osobistością statystyczną, naszym liderem i seniorem niezwykle doświadczonym i obeznanym we wszystkich sprawach statystyki.

Przez blisko 60 lat pracy w GUS Profesor przeszedł kolejne szczeble kariery zawodowej i pełnił szereg odpowiedzialnych funkcji, dzięki czemu wniósł fundamentalny wkład w rozwój i popularyzację statystyki publicznej. Był aktywnym członkiem Naukowej Rady Statystycznej, Międzynarodowego Instytutu Statystycznego, Komitetu Statystyki i Ekonometrii PAN, Rady Głównej Polskiego Towarzystwa Statystycznego oraz od 1992 r. redaktorem naczelnym miesięcznika „Wiadomości Statystyczne".

Profesor łączył pracę w statystyce z aktywnością naukową. Przez blisko 30 lat pracował w Szkole Głównej Handlowej. W dorobku Pana Profesora jest ponad 200 publikacji — książek i artykułów z zakresu wykorzystania informatyki w gospodarce, historii statystyki i historii GUS. Profesor był jednym z najwybitniejszych polskich autorytetów w dziedzinie metodologii i projektowania badań statystycznych, ochrony danych w systemach informacyjnych, zastosowań informatyki w statystyce. Jego zaangażowanie w ochronę danych osobowych mocno wspomagało obronę profesjonalnej niezależności statystyki w zakresie tajemnicy statystycznej, a publikacja dotycząca zasad projektowania i realizacji badań statystycznych służyła kolejnym generacjom młodych statystyków i porządkowała nasze myślenie oraz działanie w zakresie organizacji badań.

Imponująca była również aktywność Pana Profesora na forum międzynarodowym. Od dziesiątków lat reprezentował polską statystykę w różnych gremiach międzynarodowych, a od 1989 r. był stałym członkiem Międzynarodowego Instytutu Statystycznego. Współpracował ze statystykami światowymi w ramach wielu inicjatyw ONZ i OECD, uczestniczył także w sesjach Konferencji Statystyków Europejskich. Angażował się mocno w problematykę tajemnicy statystycznej na poziomie Europejskiego Systemu Statystycznego. Należy także podkreślić udział Pana Profesora, zarówno jako inicjatora, jak również współorganizatora i prelegenta, w wielu międzynarodowych konferencjach i seminariach naukowych.

Za ten ogromny dorobek Profesor został uhonorowany wieloma wysokimi odznaczeniami państwowymi i resortowymi, w tym Krzyżem Komandorskim Orderu Odrodzenia Polski, Krzyżem Oficerskim Orderu Odrodzenia Polski i Krzyżem Kawalerskim Orderu Odrodzenia Polski.

Dziś, przepełniony głębokim żalem, pragnę pożegnać Pana Profesora w imieniu własnym oraz społeczności polskich statystyków. Dzień pożegnania to czas zadumy i wspomnień, dla mnie szczególnie bolesny ze względu na łączącą nas przyjaźń, lata wspólnej pracy i doświadczeń związanych ze zmaganiami w realizacji ambitnych celów w zakresie rozwoju statystyki. Jestem dumny, że dane mi było zetknąć się i współpracować z Człowiekiem o tak rozległej wiedzy, pełnym

zaangażowania, inicjatywy i stałej gotowości do mierzenia się z wszelkimi przeciwnościami. Ale przede wszystkim z Człowiekiem niezwykle skromnym, którego postawa przepełniona była szacunkiem, troską i otwartością na problemy innych.

Drogi Profesorze, zapisałeś się w historii GUS jako wybitny statystyk, służący przez wiele lat swoją wiedzą i doświadczeniem. Twoje odejście to niepowetowana strata dla statystyki i wszystkich, którzy Cię znali i z Tobą współpracowali.

Żegnamy Wybitnego Naukowca, Znakomitego Statystyka, żegnamy prawego Człowieka o wielkim sercu, niezawodnie życzliwego i służącego pomocą.

Takiego Cię zapamiętamy. Będzie nam Ciebie brakowało, a Twoja nieobecność w naszej rodzinie statystycznej będzie niezwykle dotkliwa.

Dzisiaj żegnamy Cię Szanowny Profesorze i Przyjacielu,
ale pozostaniesz w naszej pamięci i w naszych sercach.

Janusz Witkowski
P r e z e s GUS

Mirosław SZREDER

Zmiany w strukturze całkowitego błędu badania próbkowego

Pragnę zadedykować niniejsze opracowanie Profesorowi Janowi Kordosowi w 85. rocznicę urodzin. Prace naukowe Jubilata stanowią dla mnie, i sądzę dla wielu innych, inspirację do głębszych dociekań nad jakością danych i jakością badań statystycznych. W tej dziedzinie Profesor Jan Kordos jest w Polsce niekwestionowanym autorytetem.

Śledząc rozwój metodyki badań statystycznych i ich praktycznego zastosowania w wielu dziedzinach życia gospodarczego i społecznego, warto zwrócić uwagę na kilka najważniejszych — moim zdaniem — zmian, jakich doświadczyliśmy w tym zakresie w ostatnich dziesięcioleciach. Po pierwsze, od połowy ub. wieku zmieniło się nieufne początkowo nastawienie wielu osób i instytucji, decydujących o sposobach badań statystycznych, do badań niewyczerpujących (próbkowych). Po drugie, istotny postęp dokonał się w rozwoju techniki próbkowania i wnioskowania, co pozwoliło nie tylko na dobrą kontrolę błędu losowania, ale także na znaczne zmniejszenie liczebności prób badawczych, a w konsekwencji czasu i kosztów realizacji badań. Obecnie, jak się wydaje, za jedną z najistotniejszych zmian w podejściu do badań próbkowych uznać należy większe niż poprzednio koncentrowanie się statystyków na tych składnikach całkowitego błędu badania, które nie mają charakteru losowego. O rosnącym znaczeniu błędów nielosowych i o próbach ich kontroli traktuje moje opracowanie.

LATA NIEUFNOŚCI DO BADAŃ NIEWYCZERPUJĄCYCH

Osiemnastowieczne idee Laplace'a na temat możliwości poznania populacji na podstawie pobranej z niej próby, podjęte następnie (ok. 1900 r.) przez norweskiego statystyka A. N. Kiaera[1] oraz brytyjskiego ekonomistę i statystyka A. L. Bowleya[2],

[1] W Wikipedii Anders Nicolai Kiaer (1838—1919) nazwany jest statystykiem, który jako pierwszy zaproponował korzystanie z reprezentatywnej próby, zamiast badania wyczerpującego, w celu zgromadzenia informacji o całej populacji (*who first proposed that a representative sample rather than a complete enumerating survey could and should be used to collect information about a population*).

[2] Arthur Lyon Bowley (1869—1957), angielski statystyk i ekonomista, jeden z pionierów wykorzystania techniki próbkowania w badaniach społecznych.

a formalnie naukowo opracowane i rozwinięte w połowie lat 30. ub. wieku przez Jerzego Spławę-Neymana (1894—1981) oraz Egona Pearsona (1895—1980), nie zyskały od razu powszechnego zaufania. Z rezerwą odnoszono się w szczególności do precyzji wnioskowania w zakresie dużych liczebnie populacji na podstawie prób stanowiących procent lub ułamek procenta całej zbiorowości.

W Stanach Zjednoczonych, w czasie gdy George Gallup z powodzeniem zaczynał wykorzystywać próbkowe badania sondażowe do przewidywania wyniku wyborczego (1936 r.), prezydent F. D. Roosevelt oraz Kongres, nieprzekonani o wiarygodności badań próbkowych, zdecydowali o przeprowadzeniu badania wyczerpującego na temat skali bezrobocia w Stanach Zjednoczonych w 1937 r. Na wysłany pocztą do gospodarstw domowych kwestionariusz dobrowolnie mieli odpowiedzieć bezrobotni. Z koncepcją tą nie zgadzało się wielu ówczesnych statystyków, podkreślając ryzyko błędów wynikających z niechęci do przyznania się części respondentów do braku pracy i wskazując na lepszy i tańszy sposób osiągnięcia tego celu poprzez badanie próbkowe. W Census Bureau zgodzono się ostatecznie na eksperymentalne przeprowadzenie badania niewyczerpującego na próbie 2% gospodarstw domowych, równolegle ze sfinansowanym przez Kongres badaniem wyczerpującym (Desrosieres, 1998). W badaniu pełnym uzyskano 7,8 mln zwrotów od osób bezrobotnych, co uznano za szacunek wielkości bezrobocia w Stanach Zjednoczonych. Tymczasem z wywiadów bezpośrednich zrealizowanych w badaniu próbkowym wynikało, że jedynie 71% faktycznie bezrobotnych wysłało wypełniony kwestionariusz. Zestawienie ze sobą tych dwu liczb oznaczało, że liczba bezrobotnych była w tym czasie bliska 11 mln, a nie 7,8 mln, co sugerowało niedoskonale zaprojektowane i wykonane badanie pełne.

Ale i współcześnie, głównie poza środowiskiem statystyków, wyrażany bywa pogląd o „niepełnowartościowych" wynikach badań reprezentacyjnych w stosunku do badań pełnych. Jeszcze kilka lat temu, gdy zaczynało przybywać w środkach masowej informacji różnego rodzaju ankiet i sondaży, niezrozumienie idei badania reprezentacyjnego i niezbędnej do jego realizacji wielkości próby było dość powszechne. Charakterystyczny dla tych postaw był w 2005 r. głos znanej i zasłużonej felietonistki „Tygodnika Powszechnego", która swoje wątpliwości (w kontekście zbliżających się wówczas wyborów prezydenckich) wyraziła następująco: *Pytam: co to jest tak naprawdę ta „próba reprezentacyjna"? Dlaczego ma ona być dla mnie ważna w jakikolwiek sposób? 1024 osoby — czy choć jedna znajduje się w kręgu tych, którzy razem ze mną jadą tramwajem, czytają te same gazety, wchodzą do tych samych sklepów? Dlaczego ta mała grupka ma być źródłem ustaleń i faktów, z którymi miałabym się liczyć? (...) 29% z nich dzisiaj głosowałoby na człowieka, który w moim własnym przekonaniu na prezydenta Rzeczypospolitej w ogóle się nie nadaje... 29% z 1024 to raptem 296 z ułamkiem. I to ma być wielkość „reprezentacyjna"; taka, która na dziś wróży zwycięstwo w skali całej Polski, dla trzydziestu kilku milionów rodaków. Czy to nie absurdalne?* (Hennelowa, 2005).

Mimo że obecnie tego typu zasadniczych wątpliwości nie obserwuje się raczej u większości dziennikarzy, to i tak stopień ich zaufania do badań próbkowych nie jest wysoki. Powodem jest, jak sądzę, niezrozumienie innego aspektu badań reprezentacyjnych, związanego właśnie z tytułową strukturą błędu, jakim mogą być obciążone wyniki wnioskowania. Powszechne jest bowiem utożsamianie błędu losowania z całkowitym błędem badania. Refleksja nad strukturą błędu w badaniach statystycznych wydaje się potrzebna także dlatego, że koncentrowanie się wyłącznie na błędzie losowania prowadzi niektórych do konstatacji, że w badaniu wyczerpującym żaden błąd nie występuje. Stąd przekonanie, że wyniki badania wyczerpującego są zawsze precyzyjniejsze od analogicznych pochodzących z badania reprezentacyjnego. Tymczasem, niezależnie od innych rodzajów błędu, same problemy zarządzania dużym badaniem wyczerpującym, takim jak spis powszechny, a także skala przetwarzania materiału liczbowego stanowić mogą źródło istotnych niedokładności ostatecznych wyników.

BŁĄD LOSOWANIA I JEGO ZNACZENIE

Trudno byłoby zaprzeczyć, że w badaniach nad błędami badań próbkowych statystycy przez dziesięciolecia koncentrowali się przede wszystkim na błędzie losowania (*sampling error*). Jest to bowiem ta kategoria błędu, która bezpośrednio wiąże się z modelem matematycznym stosowanym w próbkowaniu i wnioskowaniu, a więc łatwo zrozumiała i dość dobrze przez statystyków poznana. W przypadku braku odpowiedniego modelu (większość kategorii błędu nielosowego) znacznie trudniej jest o właściwy opis analityczny i skuteczny pomiar. Dostrzegał to już w 1951 r. słynny hinduski statystyk P. C. Mahalanobis, stwierdzając: *Ogólna postawa jest taka, że błąd nielosowy postrzega się jako coś, co nie dotyczy statystyka, a w każdym razie uważany jest za rodzaj niezbyt przyjemnej sprawy, którą dumny statystyk nie musi się przejmować*[3].

A kilkadziesiąt lat później na łamach „Wiadomości Statystycznych" przypomniano podobnie brzmiące stwierdzenie innego znakomitego statystyka Lesliego Kisha: *sampling error is „over-researched"*[4] (błąd losowania jest nadmiernie badany). W nauce sugestia o nadmiernym badaniu pewnego zagadnienia może oznaczać jedynie to, że za mało uwagi poświęca się innemu, powiązanemu z nim lub konkurencyjnemu zagadnieniu. Potrzeba podjęcia większego wysiłku

[3] *In fact, the general attitude is to look upon the non-sampling error as something, which does not concern the statistician, or in any case is a kind of dirty job, which a highbrow statistician need not bother about* (Mahalanobis, 1951), s. 4

[4] Sformułowanie to pojawiło się m.in. w artykule Richarda Platka i Carla-Erika Särndala pt. *Can a statistician deliver?* opublikowanym w języku polskim wraz z dyskusją w „Wiadomościach Statystycznych", nr 4, w 2001 r.

statystyków w odniesieniu do badań nad błędami nielosowymi jest więc tym, co łączy oba cytowane wyżej stwierdzenia.

Błąd losowania jest eksponowany w większości komentarzy do wyników badań reprezentacyjnych nie tylko dlatego, że potrafimy określić jego wielkość w konkretnym schemacie próbkowania, ale także dlatego, iż znana jest jego natura i istota. Jest to błąd, który powstaje na skutek konieczności stosowania niedoskonałej techniki próbkowania, tj. takiej, która nie gwarantuje uzyskania struktury próby w pełni zgodnej ze strukturą populacji. Gdy się natomiast przyjrzeć błędom o charakterze nielosowym, to okazuje się, że ich natura jest bardziej złożona, a zatem bardziej kłopotliwa dla statystyków. Nawet w typowej klasyfikacji odnajdujemy co najmniej cztery rodzaje błędów nielosowych, z których każdy kryje w sobie znaczne bogactwo źródeł i treści. Są to następujące błędy:

- pokrycia jednostek badanej zbiorowości przez operat (*coverage error*),
- spowodowane brakiem odpowiedzi respondentów (*nonresponse error*),
- pomiaru (*measurement error*), związane z zarejestrowaniem nieprawdziwych informacji o respondencie,
- przetwarzania danych (*postsurvey processing error*).

Dodatkowo, co warte podkreślenia, wymienionymi wyżej kategoriami błędów mogą być obciążone nie tylko wyniki badania częściowego, ale także pełnego. Specyficzny dla badań próbkowych jest jedynie błąd losowania.

Znaczenie błędu losowania wśród rozważanych kategorii błędów jest współcześnie mniejsze niż dawniej zarówno ze względu na postęp, jaki dokonał się w rozwoju teorii statystyki (w szczególności metody reprezentacyjnej), jak i z tego powodu, że relatywnie wzrosła waga błędów nielosowych w całkowitym błędzie badania. Efekty tej prawidłowości są dobrze widoczne m.in. w niekończących się dyskusjach na temat powodów różnic między wynikami sondaży politycznych wykonywanych w tym samym czasie i na ten sam temat przez różne ośrodki badawcze. Każdy z nich podaje jedynie 3% wielkość błędu losowania (nazywanego też w mediach błędem statystycznym), który nie jest w stanie wyjaśnić znacznie większych różnic w uzyskanych wynikach. Świadomość tego, że muszą istnieć efekty oddziaływania innych błędów stała się dla komentatorów jasna, gdy badania takie wykonało kilka ośrodków na większej próbie (ok. 2400 jednostek), w której błąd losowania nie przekracza 2%. Różnice w wynikach nadal się utrzymywały, co było jednym z ważnych dowodów na to, że błąd losowania jest zaledwie jednym ze składników (czasami najmniejszym) całkowitego błędu badania próbkowego.

WZROST ZNACZENIA BŁĘDÓW NIELOSOWYCH

Analizując kolejne z wymienionych wcześniej kategorii błędów nielosowych jako pierwszy pojawia się błąd pokrycia, który jest konsekwencją niedokładnego pokrycia rzeczywistej populacji badania przez użyty operat losowania. Wydawałoby się, że rola tego błędu powinna maleć w świecie, w którym mamy coraz

więcej różnych baz danych, banków danych czy rejestrów administracyjnych. Tak jednak nie jest, gdyż często kompletność i aktualność komercyjnych baz danych jest trudna do zweryfikowania, a rejestry administracyjne — nieraz dobrej jakości — rzadko mogą stanowić gotowy operat (szerzej na ten temat piszą autorzy zajmujący się tzw. *register-based statistics*). Problemy z właściwym operatem dotyczą zarówno badań prowadzonych przez podmioty statystyki publicznej[5], jak i wielu badań społecznych, w tym badań opinii publicznej. W wielu krajach jeszcze do niedawna, w czasach przed pojawieniem się telefonii komórkowej, popularnym i skutecznym sposobem losowania respondenta było generowanie losowego numeru stacjonarnego telefonu (*random digit dialing*), który identyfikował dane gospodarstwo domowe. Pojawienie się telefonów komórkowych wymusiło konieczność zmiany tej techniki, a przynajmniej włączenie do generowanych sekwencji numerów także telefonów komórkowych. Przed wyborami prezydenckimi w Stanach Zjednoczonych w 2012 r. Instytut Gallupa zmienił swój dotychczasowy sposób losowania wyborców poprzez losowe generowanie numerów telefonicznych na rzecz tańszego sposobu, polegającego na losowaniu spośród niezastrzeżonych numerów telefonów stacjonarnych (50% próby) oraz losowego generowania numerów telefonów komórkowych (pozostałe 50%). Celem tej zmiany miało być objęcie tak określonym operatem możliwie jak największej frakcji gospodarstw domowych. Jednak słabością włączenia do operatu posiadaczy telefonów komórkowych okazała się ich specyficzna struktura. Ta część próby, jak stwierdził później Instytut Gallupa[6], była wiekowo młodsza, o większej przewadze sympatii dla Partii Demokratycznej i w większym stopniu skłonna poprzeć Baracka Obamę jako prezydenta aniżeli respondenci wylosowani z racji posiadania telefonu stacjonarnego. I mimo że zastosowanie odpowiednich wag może zbliżyć tę strukturę do ogólnokrajowej, uznano to za jedno ze źródeł błędu w przewidywaniu wyniku wyborów prezydenckich (warto dodać, że błąd we wskazaniu zwycięzcy wyborów prezydenckich na podstawie prognozy wyborczej przytrafił się Instytutowi Gallupa w 2012 r. po raz pierwszy od ponad 60 lat).

Błędem, którego znaczenie rośnie w ostatnich latach najszybciej jest błąd braków odpowiedzi. Są one, z powodu dużej liczby realizowanych badań próbkowych i zmęczenia respondentów udziałem w tych badaniach, poważnym wyzwaniem dla statystyków. Niekiedy z pieczołowicie zaprojektowanego i przygotowanego próbkowania, które ma pozwolić na precyzyjne wnioskowanie statystyczne, pozostaje połowa albo i mniej oczekiwanej próby jednostek udzielających odpowiedzi. Wraz z nią pozostaje też ważne pytanie o stopień reprezenta-

[5] Wskazuje na to m.in. Paradysz (2010), s. 52, który pisze: *Wydaje się, że obecność błędów pokrycia w badaniach statystycznych GUS była dostrzegana dość dawno, ale stanowiły one problem wstydliwy, o którym oficjalnie nie pisano*.

[6] *Gallup 2012*... (2013), s. 7 i 8.

tywności zrealizowanej próby, a w kolejności, o możliwość stosowania metod i techniki wnioskowania statystycznego. Od dawna wiadomo, że brak odpowiedzi jest znaczącym ubytkiem informacyjnym w badaniu: *Brak kontaktu z respondentami lub odmowy odpowiedzi mogą poważnie obciążać dane ankietowe*[7]. *Problemem badań sondażowych jest to, że sporo osób odmawia w nich udziału, a niektórzy odmawiają odpowiedzi na poszczególne pytania. Ogranicza to wiarygodność wyników danych sondażowych*[8].

Większość projektantów i realizatorów badań, a także statystyków zajmujących się techniką uzupełniania brakujących danych podziela pogląd C. E. Särndala i S. Lundströma, że *współcześnie braki odpowiedzi są normalną (chociaż niepożądaną) cechą badań ankietowych*[9]. Procent braków odpowiedzi zależy od wielu czynników, w tym przede wszystkim od: tematyki i zakresu badania, populacji poddanej badaniu, typu pomiaru sondażowego (ankietowania) i formy pytań. Często ma też znaczenie, na ile respondent czuje się bezpieczny, jeśli chodzi o jego anonimowość (tajemnica statystyczna). Przy projektowaniu badań brane są pod uwagę wymienione czynniki w dążeniu do uzyskania jak najwyższego wskaźnika odpowiedzi respondentów w próbie. Mimo tego jednak, rzeczywistość nie jest optymistyczna. Badacze tego zagadnienia Donsbach i Traugott (2008), stwierdzają m.in.: *Z pewnymi wyjątkami wskaźniki odpowiedzi (response rates) rzadko przewyższają obecnie 50%. Nawet dobrym organizacjom badań społecznych trudno jest dzisiaj osiągnąć 50% wskaźniki odpowiedzi; najczęściej wahają się one pomiędzy 30% a 40%. Badacze rynku uznają wskaźniki 15% lub 20% za akceptowalne*. Holenderska autorka A. L. Stoop (2005) podaje z kolei przykłady ważnych badań statystycznych, w których wskaźniki braków odpowiedzi (*nonresponse rates*) wzrosły w ostatnich kilkunastu latach z 30% do 50%. W badaniach sondażowych w Polsce niewiele jest informacji na temat skali odmów. Swego rodzaju odkryciem w tym względzie stało się opracowanie ekspertów z PAN na temat przyczyn niepowodzeń sondaży przed pierwszą turą wyborów prezydenckich w 2010 r. (*Ocena…*, 2010). Autorzy wskazali tam, że w co najmniej kilku sondażach zrealizowanych przez największe instytuty badawcze w naszym kraju w czasie kampanii prezydenckiej, udziału w badaniu odmawiało ponad 80% badanych, a były też takie, w których odmowy stanowiły 93%. Na uwagę zasługuje więc i obecna skala braków odpowiedzi, jak też tendencja wzrostu wskaźnika braków odpowiedzi w czasie.

W tej ostatniej kwestii wypowiadają się interesująco Dillman, Smyth i Christan (2009). Autorzy ci z perspektywy minionych dziesięcioleci starają się uzasadnić wpływ ewolucji w używanej technice wywiadu z respondentami na ro-

[7] *Not-at-homes and refusals can seriously bias survey data* (Zikmund, 1997), s. 205.
[8] Fragment oświadczenia Organizacji Firm Badania Opinii i Rynku (OFBOR).
[9] *Today, nonresponse is a normal (but undesirable) feature of the survey undertaking* (Särndal, Lundström, 2006), s. 9.

snące wskaźniki odmów. Dokonując przeglądu metod wywiadu — od osobistego (*personal interview*) w latach 60. ub. wieku, poprzez telefoniczny w latach 70. i 80. ub. wieku, do internetowego obecnie — zwrócili oni uwagę na malejące od lat 90. ub. wieku wzajemne oddziaływanie respondenta i ankietera (*human interaction*). Jest to szczególnie widoczne, gdy porówna się współczesne ankiety internetowe z wywiadami bezpośrednimi kilku dekad XX w., kiedy ankieter był instruowany, w jaki sposób pozyskać przychylność respondenta, stworzyć atmosferę swobodnej wypowiedzi, zadbać o taki sposób wywiadu, w którym respondent odniósłby wrażenie, że ankieter jest prawdziwie zainteresowany opiniami i poglądami rozmówcy. Stopniowe słabnięcie tych więzi, gdy rozmowę bezpośrednią zastąpił telefon, a później komputer, to także coraz krótszy czas i coraz mniej uwagi poświęcanej respondentowi, a w konsekwencji także słabsze jego przekonanie o celowości (znaczeniu) badania. Rezultatem tego jest najczęściej malejąca motywacja respondenta do udziału w badaniu, które z bezpośredniej konwersacji przekształciło się w pozbawione personalnych relacji doświadczenie.

Kolejną kategorią błędu nielosowego dość często obecną w badaniach próbkowych (a także wyczerpujących) jest błąd pomiaru. Polega on na zarejestrowaniu nieprecyzyjnej lub nieprawdziwej informacji w odniesieniu do określonej cechy (zmiennej) w badaniu. Źródłem tego błędu może być zarówno ankieter, np. pośpiech i niestaranność w jego pracy, jak i respondent, który przed udzieleniem informacji wykonuje szereg myślowych operacji — od próby zrozumienia pytania, poprzez przywołanie użytecznych na ten temat informacji w swojej pamięci, ich ocenę, aż do sformułowania odpowiedzi. Podanie nieprawdziwej odpowiedzi może być zarówno intencjonalne, np. gdy respondent widzi w tym swój interes, jak i niezamierzone, spowodowane naturalną niedoskonałością ludzkiej pamięci, jak pisze Wheelan (2013): *pamięć jest rzeczą fascynującą, jednak nie zawsze bogatym źródłem dobrych danych*. Oczywiście wielkość błędu rośnie jeszcze bardziej, gdy w danym badaniu spotkają się dwie słabości respondentów — braki w pamięci (*memory failure*) i dążenie do udzielenia politycznie lub społecznie „poprawnej" odpowiedzi (*social desirability*). Dowodem tego jest wykryta w wielu krajach tendencja do deklarowania udziału w wyborach przez część respondentów, którzy w dniu głosowania pozostali w domach. The British Election Study[10], analizujące postawy brytyjskich wyborców od ponad pół wieku, ocenia, że powyborcze deklaracje co do udziału w wyborach badanych respondentów bywają aż o 20 p.proc. zawyżone. Wart przypomnienia jest w tym kontekście inny podobny błąd, polegający na przekonaniu części respondentów, że głosowali na danego polityka wówczas, gdy dobrze sprawował swój urząd. Tak było z prezydentem J. F. Kennedym, który zwyciężając w wyborach w 1960 r. z przewagą niewiele ponad 50%, zdobył w następnych latach

[10] Por. http://www.britishelectionstudy.com/ (3.11.2014).

taką popularność, że 4 lata później aż 64% amerykańskich wyborców twierdziło, iż głosowało właśnie na niego[11].

Wskazane błędy o charakterze nielosowym, mające negatywny wpływ na jakość wnioskowania, nie są jedynymi błędami, których natura jest inna od błędu losowania. Na dalszych etapach realizacji badania próbkowego (ale także wyczerpującego) pojawić się mogą błędy w przesyłaniu, przetwarzaniu lub raportowaniu danych. Krótkie omówienie najważniejszych błędów nielosowych prowadzi do konstatacji, że wraz z upowszechnianiem się badań ankietowych i sondażowych znaczenie tych błędów rośnie. Koncentrowanie się w wielu raportach z badań wyłącznie na błędzie losowania staje się współcześnie niewystarczające. Dla statystyków ważne jest jednak zarówno doskonalenie sposobów pomiaru wielkości błędów nielosowych, jak i poszukiwanie techniki minimalizowania ryzyka wystąpienia poszczególnych kategorii tych błędów w konkretnym badaniu. Sądzę, że są to jedne z ważnych wyzwań współczesnej statystyki.

prof. dr hab. Mirosław Szreder — *Uniwersytet Gdański*

LITERATURA

Desrosieres A. (1998), *The Politics of Large Numbers, A History of Statistical Reasoning*, Harvard University Press

Dillman D. A., Smyth J. D., Christan L. M. (2009), *Internet, mail and mixed-mode surveys. The tailored design method*, J. Wiley

Donsbach W., Traugott M. W. (ed.) (2008), *Public Opinion Research*, SAGE Publications

Gallup 2012 Presidential Election Polling Review (2013), Gallup, www.gallup.com

Hennelowa J. (2005), *„1024"*, „Tygodnik Powszechny", nr 22, 29 maja

Maarek P. J. (2011), *Campaign Communication and Political Marketing*, J. Wiley

Mahalanobis P. C. (1951), *Professional Training in Statistics*, „Bulletin of the International Statistical Institute", Vol. 33, No. 5

Ocena metodologii i rezultatów badań poprzedzających pierwszą i drugą turę wyborów prezydenckich 2010 roku (2010), Raport Komisji pod przewodnictwem prof. H. Domańskiego, Warszawa

Paradysz J. (2010), *Konieczność estymacji pośredniej w spisach powszechnych*, [w:] E. Gołata (red.), *Pomiar i informacja w gospodarce*, „Zeszyty Naukowe", nr 149, Uniwersytet Ekonomiczny w Poznaniu

Särndal C. E., Lundström S. (2006), *Estimation in Surveys with Nonresponse*, J. Wiley

Stoop A. L. (2005), *The Hunt for the Last Respondent*, Social and Cultural Planning Office of the Netherlands, The Hague

Wheelan Ch. (2013), *Naked statistics. Stripping the Dread from the Data*, W. W. Norton and Co.

Zikmund W. G. (1997), *Business Research Methods* (5th ed.), The Dryden Press

[11] Maarek (2011).

SUMMARY

The author attempts to justify the thesis on an increase in the error weight of a non-random character in the total error of sample testing. In particular, he focuses on the non-response, coverage and measurement errors. The growing importance of these errors in all kinds of surveys and sounding surveys cause that insufficient becomes identifying the messages of these studies only draw the error and its size. This article is dedicated to Professor Jan Kordos, the skilled and authority in the field of quality of statistical data on the occasion of 85th anniversary of his birth.

РЕЗЮМЕ

Автор старается обосновать тезис по росте веса ошибок с неслучайным характером в полной ошибке выборочного обследования. В частности его внимание сосредоточивается на ошибках: отсутствия ответов, охвата и измерения. Увеличение значения этих ошибок в разных анкетных и зондажных обследованиях показывает, что недостаточным является указание в статистических сообщениях на ошибку выборки и ее размер. Автор посвятил эту статью Профессору Яну Кордосу, специалисту и научному авторитету в области обследований качества статистических данных по случаю 85-летия со дня Его рождения.

Piotr SULEWSKI

Miary związku między cechami w tablicy trójdzielczej

Metody wnioskowania dotyczące jednej cechy są bardzo popularne w literaturze statystycznej. Jednak w badaniach statystycznych bardzo często obiekty opisuje się za pomocą więcej niż jednej cechy. Bardzo popularne narzędzie do badania związku między cechami stanowią tablice wielodzielcze. Jeżeli dane obejmują dwie cechy, to mówimy o tablicach dwudzielczych, trzy cechy — tablicach trójdzielczych, cztery cechy — tablicach czterodzielczych, ..., *m* cech — tablicach *m*-dzielczych (Oktaba, 1974). Tablicami wielodzielczymi są także kostki OLAP stosowane np. w Narodowym Spisie Ludności i Mieszkań czy w Powszechnym Spisie Rolnym.

W publikacji Sulewskiego (2007) opisano procedurę generowania tablic dwudzielczych o zadanym związku między cechami. Zbadano także moc testu, czyli zdolność tablicy dwudzielczej do wykrywania związku między badanymi cechami oraz obliczono jego siłę wykorzystując współczynnik V Cramera. Porównano rzeczywisty rozkład statystyki testowej dla tablic dwudzielczych z rozkładem teoretycznym, jakim jest rozkład chi-kwadrat o $(w-1)(k-1)$ stopniach swobody. Dzięki tej publikacji czytelnik może samodzielnie modelować tablice o rozmiarach, z którymi ma do czynienia w praktyce i może badać ich moc.

W innym opracowaniu (Sulewski, 2013) przedstawiono test niezależności chi-kwadrat dla tablicy trójdzielczej wraz z jego implementacją komputerową w języku VBA. Zbadano też zdolność tablicy trójdzielczej do wykrywania związku między badanymi cechami.

Wśród użytkowników komputerów bardzo popularne są arkusze kalkulacyjne, a w szczególności Microsoft Excel. Do zalet tej aplikacji zaliczamy m.in.: łatwe gromadzenie danych i na ich podstawie tworzenie tabel oraz wykresów, stosowanie różnorodnych obliczeń z zastosowaniem funkcji matematycznych i statystycznych, a także możliwość zwiększania operatywności programu za pomocą kodów VBA (*Visual Basic for Applications*).

W literaturze statystycznej bardzo popularne są narzędzia do pomiaru siły związku między cechami w tablicach dwudzielczych. Należą do nich m.in. współczynniki: V Cramera, T Czuprowa, kontyngencji C Pearsona, φ Yule'a, Goodmana-Kruskala (Goodman, Kruskal, 1954), κ Cohena (Cohen, 1960), Spearmana, *eta* (*SPSS*), współczynnik punktowo-dwuseryjny (Sobczyk, 1996) i niepewności (*SPSS*). Jednak mechanizmy do pomiaru siły związku między cechami w tablicach trójdzielczych nie są już tak bardzo popularne, a w literaturze polskojęzycznej praktycznie nie występują.

W artykule zdefiniowano tablicę trójdzielczą wraz z jej charakterystykami oraz przedstawiono implementację komputerową dotyczącą pomiaru siły związku między trzema cechami za pomocą współczynników: Gray-Williamsa, Marcotorchino i Lombardo.

Pierwszym celem pracy jest przedstawienie teorii dotyczącej tablic trójdzielczych. Jest to niezbędne do lepszego zrozumienia zagadnień dotyczących pomiaru siły związku między trzema cechami. Drugim celem jest przedstawienie gotowych procedur i funkcji zapisanych w języku VBA, umożliwiających czytelnikowi samodzielne przeprowadzanie badań statystycznych w najbardziej popularnym arkuszu kalkulacyjnym, jakim jest Microsoft Excel.

TABLICA TRÓJDZIELCZA I JEJ CHARAKTERYSTYKI

Tablicę, która gromadzi wyniki podziału próby według trzech cech X, Y, Z nazywamy tablicą trójdzielczą. Można sobie ją wyobrazić jako kostkę z *„w”* wierszami, *„k”* kolumnami i *„p”* płaszczyznami. Tablicę trójdzielczą można także zilustrować rozkładając płaszczyzny *„p”* obok siebie (tabl. 1 i 2).

Gdy liczba wszystkich przyjmowanych przez cechy X, Y, Z wartości jest względnie mała, wpisuje się je w odpowiednie wiersze, kolumny i płaszczyzny. Z kolei gdy liczba możliwych wartości jest duża niezbędne okazuje się ich pogrupowanie w klasy.

Tablica trójdzielcza ułatwia odczytywanie zależności między cechami tak ilościowymi, jak i jakościowymi. Najprostszą jej postacią jest tablica $2\times2\times2$ (tabl. 1), która składa się z 8 liczebności n_{ijt} $(i, j, t = 1, 2)$ rozkładu łącznego cech X, Y, Z.

TABL. 1. TABLICA TRÓJDZIELCZA 2×2×2

Wyszczególnienie	Z_1		Z_2		Razem
	Y_1	Y_2	Y_1	Y_2	
X_1	n_{111}	n_{121}	n_{112}	n_{122}	$n_{1\bullet\bullet}$
X_2	n_{211}	n_{221}	n_{212}	n_{222}	$n_{2\bullet\bullet}$
Razem	$n_{\bullet11}$	$n_{\bullet21}$	$n_{\bullet12}$	$n_{\bullet22}$	n

Ź r ó d ł o: opracowanie własne.

Sumy brzegowe w tablicy trójdzielczej $2\times2\times2$ to:

$$n_{11\bullet} = n_{111} + n_{112},\ n_{12\bullet} = n_{121} + n_{122},\ n_{21\bullet} = n_{211} + n_{212},\ n_{22\bullet} = n_{221} + n_{222} \quad (1)$$

$$n_{1\bullet1} = n_{111} + n_{121},\ n_{1\bullet2} = n_{112} + n_{122},\ n_{2\bullet1} = n_{211} + n_{221},\ n_{2\bullet2} = n_{212} + n_{222} \quad (2)$$

$$n_{\bullet11} = n_{111} + n_{211},\ n_{\bullet12} = n_{112} + n_{212},\ n_{\bullet21} = n_{121} + n_{221},\ n_{\bullet22} = n_{122} + n_{222} \quad (3)$$

$$n_{1\bullet\bullet} = n_{111} + n_{112} + n_{121} + n_{122} = n_{11\bullet} + n_{12\bullet} \tag{4}$$

$$n_{2\bullet\bullet} = n_{211} + n_{212} + n_{221} + n_{222} = n_{21\bullet} + n_{22\bullet} \tag{5}$$

$$n_{\bullet 1\bullet} = n_{111} + n_{112} + n_{211} + n_{212} = n_{11\bullet} + n_{21\bullet} \tag{6}$$

$$n_{\bullet 2\bullet} = n_{121} + n_{122} + n_{221} + n_{222} = n_{12\bullet} + n_{22\bullet} \tag{7}$$

$$n_{\bullet\bullet 1} = n_{111} + n_{121} + n_{211} + n_{221} = n_{1\bullet 1} + n_{2\bullet 1} \tag{8}$$

$$n_{\bullet\bullet 2} = n_{112} + n_{122} + n_{212} + n_{222} = n_{1\bullet 2} + n_{2\bullet 2} \tag{9}$$

Tablicę trójdzielczą $w \times k \times p$, która składa się z $w \cdot k \cdot p$ liczebności n_{ijt} ($i = 1, ..., w$; $j = 1, ..., k$; $t = 1, ..., p$) rozkładu łącznego cech X, Y, Z zaprezentowano w tabl. 2.

TABL. 2. TABLICA TRÓJDZIELCZA $w \times k \times p$

Wyszczególnienie	Z_1				...	Z_p				Razem
	Y_1	Y_2	...	Y_k	...	Y_1	Y_2	...	Y_k	
X_1	n_{111}	n_{121}	⋯	n_{1k1}	⋯	n_{11p}	n_{12p}	⋯	n_{1kp}	$n_{1\bullet\bullet}$
X_2	n_{211}	n_{221}	⋯	n_{2k1}	⋯	n_{21p}	n_{22p}	⋯	n_{2kp}	$n_{2\bullet\bullet}$
⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
X_w	n_{w11}	n_{w21}	⋯	n_{wk1}	⋯	n_{w1p}	n_{w2p}	⋯	n_{wkp}	$n_{w\bullet\bullet}$
Razem	$n_{\bullet 11}$	$n_{\bullet 21}$	⋯	$n_{\bullet k1}$	⋯	$n_{\bullet 1p}$	$n_{\bullet 2p}$	⋯	$n_{\bullet kp}$	n

Ź r ó d ł o: jak przy tabl. 1.

Sumy brzegowe w tablicy trójdzielczej $w \times k \times p$ to:

$$n_{11\bullet} = \sum_{t=1}^{p} n_{11t},\ n_{12\bullet} = \sum_{t=1}^{p} n_{12t},\ ...,\ n_{wk\bullet} \sum_{t=1}^{p} n_{wkt} \tag{10}$$

$$n_{1\bullet 1} = \sum_{j=1}^{k} n_{1k1},\ n_{1\bullet 2} = \sum_{j=1}^{k} n_{1k2},\ ...,\ n_{w\bullet p} \sum_{j=1}^{k} n_{wkp} \tag{11}$$

$$n_{\bullet 11} = \sum_{i=1}^{w} n_{w11},\ n_{\bullet 12} = \sum_{i=1}^{w} n_{w12},\ ...,\ n_{\bullet kp} \sum_{i=1}^{w} n_{wkp} \tag{12}$$

$$n_{1\bullet\bullet} = \sum_{j=1}^{k}\sum_{t=1}^{p} n_{1jt},\ n_{2\bullet\bullet} \sum_{j=1}^{k}\sum_{t=1}^{p} n_{2jt},\ \ldots,\ n_{w\bullet\bullet} \sum_{j=1}^{k}\sum_{t=1}^{p} n_{wjt} \tag{13}$$

$$n_{\bullet 1\bullet} = \sum_{i=1}^{w}\sum_{t=1}^{p} n_{i1t},\ n_{\bullet 2\bullet} \sum_{i=1}^{w}\sum_{t=1}^{p} n_{i2t},\ \ldots,\ n_{\bullet k\bullet} \sum_{i=1}^{w}\sum_{t=1}^{p} n_{ikt} \tag{14}$$

$$n_{\bullet\bullet 1} = \sum_{i=1}^{w}\sum_{j=1}^{k} n_{ij1},\ n_{\bullet\bullet 2} \sum_{i=1}^{w}\sum_{j=1}^{k} n_{ij2},\ \ldots,\ n_{\bullet\bullet p} \sum_{i=1}^{w}\sum_{j=1}^{k} n_{ijp} \tag{15}$$

Wartość n jest sumą wszystkich liczebności tablicy trójdzielczej:

$$\begin{aligned} n &= \sum_{i=1}^{w} n_{i\bullet\bullet} = \sum_{j=1}^{k} n_{\bullet j\bullet} = \sum_{t=1}^{p} n_{\bullet\bullet t} = \sum_{i=1}^{w}\sum_{j=1}^{k} n_{ij\bullet} = \\ &= \sum_{i=1}^{w}\sum_{t=1}^{p} n_{i\bullet t} = \sum_{j=1}^{k}\sum_{t=1}^{p} n_{\bullet jt} = \sum_{i=1}^{w}\sum_{j=1}^{k}\sum_{t=1}^{p} n_{ijt} \end{aligned} \tag{16}$$

Jeżeli cechy X, Y, Z są cechami ciągłymi, to wartości X_i, Y_j, Z_t w tablicy trójdzielczej są środkami przedziałów.

Wprowadzenie pojęcia liczebności oczekiwanej oraz liczebności resztowych ułatwi formalizację zapisu tablicy trójdzielczej.

Liczebności oczekiwane i resztowe

Niech p_{ijt} ($i = 1, 2, \ldots, w$; $j = 1, 2, \ldots, k$; $t = 1, 2, \ldots, p$) będzie prawdopodobieństwem łącznym wystąpienia zdarzenia polegającego na tym, że dany element próby będzie należał do i-tej kategorii według cechy X, do j-tej kategorii według cechy Y oraz do t-tej kategorii według cechy Z. Niezależność cech X, Y, Z oznacza, że dla wszystkich trójek wskaźników i, j, t zachodzi równość:

$$p_{ijt} = p_{i\bullet\bullet} \cdot p_{\bullet j\bullet} \cdot p_{\bullet\bullet t} \tag{17}$$

gdzie:

$p_{i\bullet\bullet}$ — prawdopodobieństwo, że obserwacja należy do i-tej kategorii według cechy X;

$p_{\bullet j\bullet}$ — prawdopodobieństwo, że obserwacja należy do j-tej kategorii według cechy Y;

$p_{\bullet\bullet t}$ — prawdopodobieństwo, że obserwacja należy do t-tej kategorii według cechy Z.

Nieznane prawdopodobieństwa p_{ijt}, $p_{i\bullet\bullet}$, $p_{\bullet j\bullet}$, $p_{\bullet\bullet t}$ oblicza się z następujących relacji:

$$n_{ijt} = n \cdot p_{ijt}, \quad n_{i\bullet\bullet} = n \cdot p_{i\bullet\bullet}, \quad n_{\bullet j\bullet} = n \cdot p_{\bullet j\bullet}, \quad n_{\bullet\bullet t} = n \cdot p_{\bullet\bullet t} \tag{18}$$

gdzie:

n — liczebność próby,
n_{ijt} — liczebności komórek,
$n_{i\bullet\bullet}$ — sumy brzegowe w wierszach,
$n_{\bullet j\bullet}$ — sumy brzegowe w kolumnach.

Liczebności oczekiwane wyznacza się ze wzoru:

$$e_{ijt} = n \cdot p_{i\bullet\bullet} \cdot p_{\bullet j\bullet} \cdot p_{\bullet\bullet t} \tag{19}$$

Stąd ostatecznie na mocy (18):

$$e_{ijt} = \frac{n_{i\bullet\bullet} \cdot n_{\bullet j\bullet} \cdot n_{\bullet\bullet t}}{n^2} \quad (i = 1, 2, ..., w;\ j = 1, 2, ..., k;\ t = 1, 2, ..., p) \tag{20}$$

Sumy brzegowe liczebności oczekiwanych e_{ijt} w tablicy $w \times k \times p$ są takie same, jak sumy brzegowe liczebności n_{ijt}. W szczególności sumy brzegowe dla wierszy w tablicy $2 \times 2 \times 2$ wynoszą:

$$\begin{aligned} e_{111} + e_{112} + e_{121} + e_{122} &= \frac{n_{1\bullet\bullet} \cdot n_{\bullet 1\bullet} \cdot n_{\bullet\bullet 1}}{n^2} + \frac{n_{1\bullet\bullet} \cdot n_{\bullet 1\bullet} \cdot n_{\bullet\bullet 2}}{n^2} + \\ &+ \frac{n_{1\bullet\bullet} \cdot n_{\bullet 2\bullet} \cdot n_{\bullet\bullet 1}}{n^2} + \frac{n_{1\bullet\bullet} \cdot n_{\bullet 2\bullet} \cdot n_{\bullet\bullet 2}}{n^2} = \\ &= \frac{n_{1\bullet\bullet}(n_{\bullet 1\bullet} \cdot n_{\bullet\bullet 1} + n_{\bullet 1\bullet} \cdot n_{\bullet\bullet 2} + n_{\bullet 2\bullet} \cdot n_{\bullet\bullet 1} + n_{\bullet 2\bullet} \cdot n_{\bullet\bullet 2})}{n^2} = \\ &= \frac{n_{1\bullet\bullet}[n_{\bullet 1\bullet}(n_{\bullet\bullet 1} + n_{\bullet\bullet 2}) + n_{\bullet 2\bullet}(n_{\bullet\bullet 1} + n_{\bullet\bullet 2})]}{n^2} = \\ &= \frac{n_{1\bullet\bullet}[n_{\bullet 1\bullet} \cdot n + n_{\bullet 2\bullet} \cdot n]}{n^2} = \frac{n \cdot n_{1\bullet\bullet}[n_{\bullet 1\bullet} + n_{\bullet 2\bullet}]}{n^2} = n_{1\bullet\bullet} \end{aligned} \tag{21}$$

$$\begin{aligned} e_{211} + e_{212} + e_{221} + e_{222} &= \frac{n_{2\bullet\bullet} \cdot n_{\bullet 1\bullet} \cdot n_{\bullet\bullet 1}}{n^2} + \frac{n_{2\bullet\bullet} \cdot n_{\bullet 1\bullet} \cdot n_{\bullet\bullet 2}}{n^2} + \\ &+ \frac{n_{2\bullet\bullet} \cdot n_{\bullet 2\bullet} \cdot n_{\bullet\bullet 1}}{n^2} + \frac{n_{2\bullet\bullet} \cdot n_{\bullet 2\bullet} \cdot n_{\bullet\bullet 2}}{n^2} = ... = \\ &= \frac{n_{2\bullet\bullet}[n_{\bullet 1\bullet} \cdot n + n_{\bullet 2\bullet} \cdot n]}{n^2} = \frac{n \cdot n_{2\bullet\bullet}[n_{\bullet 1\bullet} + n_{\bullet 2\bullet}]}{n^2} = n_{2\bullet\bullet} \end{aligned} \tag{22}$$

Liczebności resztowe wyznacza się ze wzoru:

$$R_{ijt} = n_{ijt} - e_{ijt} \quad (i = 1, 2, ..., w;\ j = 1, 2, ..., k;\ t = 1, 2, ...,\ p) \tag{23}$$

Liczebności resztowe standaryzowane wyznacza się ze wzoru:

$$RN_{ijt} = \frac{n_{ijt} - e_{ijt}}{\sqrt{e_{ijt}}} \quad (i = 1, 2, ..., w;\ j = 1, 2, ..., k;\ t = 1, 2, ...,\ p) \tag{24}$$

Liczebności resztowe skorygowane wyznacza się ze wzoru:

$$RS_{ijt} = \frac{n_{ijt} - e_{ijt}}{\sqrt{e_{ijt}\left(1 - \frac{n_{i\bullet\bullet}}{n}\right)\left(1 - \frac{n_{\bullet j\bullet}}{n}\right)\left(1 - \frac{n_{\bullet\bullet t}}{n}\right)}} \ (i = 1, 2, ..., w;\ j = 1, 2, ..., k;\ t = 1, 2, ..., p) \tag{25}$$

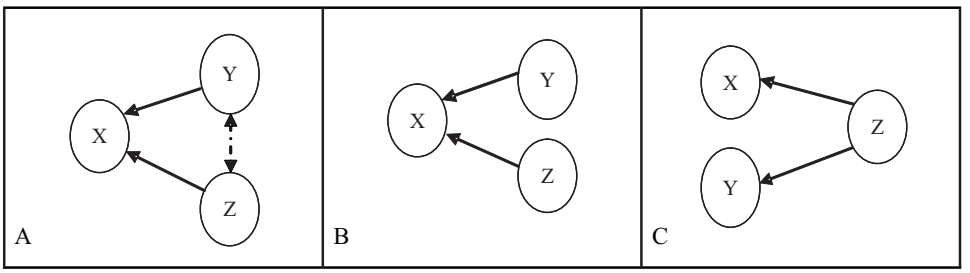
WSPÓŁCZYNNIKI KORELACJI TRZECH CECH

Współczynnik τ Goodmana-Kruskala jest popularnym współczynnikiem korelacji cech porządkowych dla tablic dwudzielczych. Numerycznymi rozszerzeniami współczynnika τ Goodmana-Kruskala dla tablic trójdzielczych są: współczynnik Gray-Williamsa, współczynnik Marcotorchino oraz współczynnik Lombardo.

Współczynnik Gray-Williamsa stosowany jest w sytuacji, gdy dwie cechy (np. *Y*, *Z*) są cechami objaśniającymi lub predyktorami wzajemnie zależnymi (*dependent predictor variables*), a trzecia cecha (np. *X*) jest cechą objaśnianą lub odpowiedzi (*response variable*) (wykr. 1A).

Współczynnik Marcotorchino stosowany jest w sytuacji, gdy dwie cechy (np. *Y*, *Z*) są cechami objaśniającymi lub predyktorami wzajemnie niezależnymi (*independent predictor variables*), a trzecia cecha (np. *X*) jest cechą objaśnianą lub odpowiedzi (*response variable*) (wykr. 1B).

Współczynnik Lombardo stosowany jest w sytuacji, gdy dwie cechy (np. *X*, *Y*) są cechami objaśnianymi lub odpowiedzi (*independent response variables*), a trzecia cecha (np. *Z*) jest cechą objaśniającą lub predyktorem (*predictor variable*) (wykr. 1C).



Ź r ó d ł o: opracowanie własne.

Współczynnik Gray-Williamsa

Jeżeli Y, Z są cechami objaśniającymi wzajemnie zależnymi, a X jest cechą objaśnianą, to niesymetryczny współczynnik Gray-Williamsa (Gray, Williams, 1975) ma postać:

$$\tau_{GW} = \frac{\sum_{i=1}^{w}\sum_{j=1}^{k}\sum_{t=1}^{p} p_{\bullet jt}\left(\frac{p_{ijt}}{p_{\bullet jt}} - p_{i\bullet\bullet}\right)}{1-\sum_{i=1}^{w} p_{i\bullet\bullet}^2} \tag{26}$$

gdzie $p_x = n_x/n$.

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

Na podstawie próby n-elementowej pobranej z populacji generalnej zweryfikowano hipotezę zerową, że cechy Y i Z nie wpływają istotnie na cechę X. Jeżeli hipoteza zerowa jest prawdziwa, to statystyka (Light, Margolin, 1971)

$$C_{GW} = (n-1)(w-1)\tau_{GW} \tag{27}$$

ma asymptotyczny rozkład chi-kwadrat z Sw_{GW} stopniami swobody, tj.:

$$Sw_{GW} = (w-1)(k-1)+(w-1)(p-1)+(w-1)(k-1)(p-1) \tag{28}$$

Wartość *p-value* wyraża się wzorem:

$$pv_{GW} = 1 - F(C_{GW}, Sw_{GW}) \tag{29}$$

gdzie $F(.)$ — dystrybuanta rozkładu chi-kwadrat dana wzorem:

$$F(x, k) = \frac{\Gamma_n(k/2, x/2)}{\Gamma(k/2)} \tag{30}$$

$\Gamma_n(.)$ jest niepełną funkcją gamma daną wzorem:

$$\Gamma_n(c, x) = \int_0^x u^{c-1} \exp(-u) du \tag{31}$$

natomiast $\Gamma(.)$ jest funkcją gamma daną wzorem:

$$\Gamma(c) = \int_0^\infty u^{c-1} \exp(-u) du \tag{32}$$

Jeżeli na poziomie istotności α mamy $pv_{GW} \leq \alpha$, to hipoteza H_0 zostaje odrzucona, gdy natomiast $pv_{GW} > \alpha$, to nie ma podstaw do odrzucenia H_0.

Współczynnik Marcotorchino

Przyjmując, że *Y*, *Z* są cechami objaśniającymi wzajemnie niezależnymi, a *X* jest cechą objaśnianą, to niesymetryczny współczynnik Marcotorchino ma postać (Marcotorchino, 1984):

$$\tau_M = \frac{\sum_{i=1}^{w}\sum_{j=1}^{k}\sum_{t=1}^{p} p_{\bullet j\bullet} \cdot p_{\bullet\bullet t}\left(\frac{p_{ijt}}{p_{\bullet j\bullet} \cdot p_{\bullet\bullet t}} - p_{i\bullet\bullet}\right)^2}{1 - \sum_{i=1}^{w} p_{i\bullet\bullet}^2} \tag{33}$$

gdzie $p_x = n_x/n$. Zakładając, że $p_{ijt} = p_{i\bullet\bullet} p_{\bullet j\bullet} p_{\bullet\bullet t}$, to $\tau_M = 0$, natomiast $p_{ijt} = p_{i\bullet\bullet} p_{\bullet\bullet t}$, wtedy $\tau_M = 1$.

Na podstawie próby *n*-elementowej pobranej z populacji generalnej zweryfikowano hipotezę zerową, mówiącą że cechy *Y* i *Z* nie wpływają istotnie na cechę *X*. Przyjmując, że hipoteza zerowa jest prawdziwa, to statystyka (Beh i in., 2007)

$$C_M = (n-1)(w-1)\tau_M \tag{34}$$

ma asymptotyczny rozkład chi-kwadrat z Sw_M stopniami swobody, gdzie:

$$Sw_M = (w-1)(k-1)+(w-1)(p-1)+(k-1)(p-1)+(w-1)(k-1)(p-1) \tag{35}$$

Wartość *p-value* wyraża się wzorem:

$$pv_M = 1 - F(C_M, Sw_M) \tag{36}$$

gdzie $F(.)$ — dystrybuanta rozkładu chi-kwadrat dana wzorem (30). Jeżeli na poziomie istotności α mamy $pv_M \leq \alpha$, to hipoteza H_0 zostaje odrzucona, ale gdy $pv_M > \alpha$, wtedy nie ma podstaw do odrzucenia H_0.

Współczynnik Lombardo

Zakładając, że *X*, *Y* są cechami objaśnianymi, a *Z* jest cechą objaśniającą, to niesymetryczny współczynnik Lombardo (współczynnik delta) ma postać (Lombardo, 2011):

$$\tau_L = \frac{\sum_{i=1}^{w}\sum_{j=1}^{k}\sum_{t=1}^{p} p_{\bullet\bullet t}\left(\frac{p_{ijt}}{p_{\bullet\bullet t}} - p_{i\bullet\bullet}\cdot p_{\bullet j\bullet}\right)^2}{1-\sum_{i=1}^{w}\sum_{j=1}^{k} p_{i\bullet\bullet}^2 \cdot p_{\bullet j\bullet}^2} \tag{37}$$

gdzie $p_x = n_x/n$. Jeżeli $p_{ijt} = p_{i\bullet\bullet}p_{\bullet j\bullet}p_{\bullet\bullet t}$, to $\tau_L = 0$. Na przykład dla tablicy $2\times2\times2$ wartość maksymalna $\tau_L = 1$, gdy $p_{1jt} = 0{,}5$ dla pewnych $(j, t = 1, 2)$ i $p_{2bc} = 0{,}5$ dla pewnych $(b, c = 1, 2; b \neq j \wedge c \neq t)$.

Na podstawie próby *n*-elementowej pobranej z populacji generalnej zweryfikowano hipotezę zerową, że cecha *Z* nie wpływa istotnie na cechy *X* i *Y*. Jeżeli hipoteza zerowa jest prawdziwa, to statystyka (Lombardo, 2011)

$$C_L = (n-1)(w-1)\ \tau_L \tag{38}$$

ma asymptotyczny rozkład chi-kwadrat z Sw_L stopniami swobody, gdzie:

$$Sw_L = (w-1)(k-1)+(w-1)(p-1)+(k-1)(p-1)+(w-1)(k-1)(p-1) \tag{39}$$

Wartość *p-value* wyraża się wzorem:

$$pv_L = 1 - FC(C_L, Sw_L) \tag{40}$$

gdzie $F(.)$ jest dystrybuantą rozkładu chi-kwadrat daną wzorem (30). Jeżeli na poziomie istotności α mamy $pv_L \leq \alpha$, to hipoteza H_0 zostaje odrzucona, jeżeli $pv_L > \alpha$, to nie ma podstaw do odrzucenia H_0.

IMPLEMENTACJA KOMPUTEROWA

Implementację komputerową pomiaru siły związku między cechami w tablicy trójdzielczej utworzoną w edytorze VBA arkusza kalkulacyjnego Excel przedstawiono w pliku „mierniki3d.xls", który umieszczono w Internecie pod adresem http://www.utogim.eu/mierniki3d.xls. Składają się na nią procedury i funkcje użytkownika umieszczone w zestawieniu (1). Aby ułatwić czytelnikowi zrozumienie użytych kodów wzbogacono je o liczne komentarze umieszczone po apostrofach i zapisane kursywą, a ponadto w zestawieniu (2) zawarto opis zmiennych wykorzystanych przy tworzeniu implementacji komputerowej.

ZESTAWIENIE (1) PROCEDUR I FUNKCJI TWORZĄCYCH IMPLEMENTACJĘ KOMPUTEROWĄ

Wyszczególnienie	Realizowane zadanie
Procedury	
WspKor3D	wyznaczanie współczynników korelacji cech *X*, *Y*, *Z*
Gray-Williams	obliczanie wartości współczynnika Gray-Williamsa
Marcotorchino	obliczanie wartości współczynnika Marcotorchino
Lombardo	obliczanie wartości współczynnika Lombardo
Funkcje	
Rozmiar3	ustalanie liczby kategorii cech *X*, *Y*, *Z*
GPijt	wyznaczanie wartości prawdopodobieństw p_{ijt}
SX3	obliczanie prawdopodobieństw brzegowych $p_{i\bullet}$
SY3	obliczanie prawdopodobieństw brzegowych $p_{\bullet j\bullet}$
SZ3	obliczanie prawdopodobieństw brzegowych $p_{\bullet\bullet t}$
SYZ3	obliczanie prawdopodobieństw brzegowych $p_{\bullet jt}$
TestChiKw	realizacja testu chi-kwadrat

Ź r ó d ł o: jak przy tabl. 1.

ZESTAWIENIE (2) ZMIENNYCH WYKORZYSTYWANYCH W IMPLEMENTACJI KOMPUTEROWEJ

Zmienne	Opis zmiennej
w, k, p	liczba kategorii w tablicy trójdzielczej
pikk	sumy brzegowe prawdopodobieństw p_i
pkjk	sumy brzegowe prawdopodobieństw p
pkkt	sumy brzegowe prawdopodobieństw p_t
pijt	wartości prawdopodobieństw p_{ijt}

ZESTAWIENIE (2) ZMIENNYCH WYKORZYSTYWANYCH W IMPLEMENTACJI KOMPUTEROWEJ (dok.)

Zmienne	Opis zmiennej
TauGw	wartość współczynnika Gray-Williamsa
TauM	wartość współczynnika Marcotorchino
TauL	wartość współczynnika Lombardo
SwGw	liczba stopni swobody dla współczynnika Gray-Williamsa
SwM	liczba stopni swobody dla współczynnika Marcotorchino
SwL	liczba stopni swobody dla współczynnika Lombardo
i, j, t, s	zmienne pomocnicze
Wynik	okno dialogowe z wynikami
n	liczebność próby
alfa	poziom istotności
Ho	hipoteza zerowa
CGw	statystyka testowa dla współczynnika Gray-Williamsa
CM	statystyka testowa dla współczynnika Marcotorchino
CL	statystyka testowa dla współczynnika Lombardo
PvGw	wartość *p-value* dla współczynnika Gray-Williamsa
PvM	wartość *p-value* dla współczynnika Marcotorchino
PvL	wartość *p-value* dla współczynnika Lombardo

Ź r ó d ł o: jak przy tabl. 1.

Przykład 1

Zbadano wpływ miejsca pochodzenia (Z) i opieki (Y) (zmienne objaśniające zależne) na diagnozę dotyczącą choroby wątroby (X) (zmienna objaśniana). Otrzymane wyniki zestawiono w formie tablicy trójdzielczej $4\times2\times3$ (tabl. 3).

TABL. 3. TABLICA TRÓJDZIELCZA 4×2×3

Diagnozy chorób wątroby	Miejsce pochodzenia diagnozy						
	północ		środek		południe		razem
	opieka medyczna						
	tak	nie	tak	nie	tak	nie	
Bezobjawowa	1	10	0	10	0	11	32
Zapalenie	43	64	31	29	89	66	322
Marskość	4	25	6	7	13	39	94
Zakażenie HCC	1	5	0	1	0	4	11
Razem	49	104	37	47	102	120	459

Ź r ó d ł o: Lombardo (2011).

Po utworzeniu tablicy trójdzielczej (tabl. 3) w arkuszu „tablica" (począwszy od komórki B2) uruchomiono procedurę „WspKor3D". W arkuszu „wyniki" za pomocą pola kombi wybrano opcję „Gray-Williamsa" i wyznaczono wartość współczynnika Gray-Williamsa, gdy Y, Z są cechami objaśniającymi wzajemnie zależnymi, a X jest cechą objaśnianą. Następnie założono hipotezę zerową, że pocho-

dzenie (*Z*) i opieka (*Y*) nie wpływają istotnie na diagnozę dotyczącą choroby wątroby (*X*). Wprowadzono poziom istotności testu α. Skoro $pv_{GW} = 0{,}000 \leq \alpha$, zatem są podstawy do odrzucenia hipotezy zerowej, czyli pochodzenie i opieka wpływają istotnie na diagnozę dotyczącą choroby wątroby (zestawienie 3).

ZESTAWIENIE (3) WYNIKÓW IMPLEMENTACJI KOMPUTEROWEJ DLA WSPÓŁCZYNNIKA GRAY-WILLIAMSA

Mierniki statystyczne	Charakterystyka współczynnika Gray-Wiliamsa
Wartość współczynnika	0,069
Poziom istotności	0,05
Hipoteza zerowa H_0	*Y* i *Z* nie wpływają istotnie na *X*
p-value	0,000
Wartość statystyki	94,41
Liczba stopni swobody	15
Są podstawy do odrzucenia hipotezy zerowej	

Ź r ó d ł o: jak przy tabl. 1.

Przykład 2

Zbadano wpływ urbanizacji (*Z*) i położenia (*Y*) (zmienne objaśniające niezależne) na preferencję oliwek (*X*) (zmienna objaśniana) wyrażoną w sześciopunktowej skali porządkowej (A—F). Otrzymane wyniki zestawiono w formie tablicy trójdzielczej $6 \times 3 \times 2$ (tabl. 4).

TABL. 4. TABLICA TRÓJDZIELCZA 6×3×2

Wyszczególnienie	Preferencja oliwek przez zamieszkujących						
	miasto			wieś			razem
	NW	*NE*	*SW*	*NW*	*NE*	*SW*	
A	20	18	12	30	23	11	114
B	15	17	9	22	18	9	90
C	12	18	23	21	20	26	120
D	17	18	21	17	18	19	110
E	16	6	19	8	10	17	76
F	28	25	30	12	15	24	134
Razem	108	102	114	110	104	106	644

Ź r ó d ł o: jak przy tabl. 3.

Po utworzeniu tablicy trójdzielczej (tabl. 4) w arkuszu „tablica” (począwszy od komórki B2) uruchomiono procedurę „WspKor3D”. W arkuszu „wyniki” za pomocą pola kombi wybrano opcję „Marcotorchino” i wyznaczono wartość współczynnika Marcotorchino, gdy *Y* i *Z* są cechami objaśniającymi wzajem-

nie niezależnymi, a X jest cechą objaśnianą. Następnie założono hipotezę zerową, że cechy urbanizacja (Z) i położenie (Y) nie wpływają istotnie na preferencję oliwek (X). Następnie wprowadzono poziom istotności testu α. Ponieważ $pv_M = 0{,}004 \leq \alpha$, zatem są podstawy do odrzucenia hipotezy zerowej, czyli urbanizacja i położenie wpływają istotnie na preferencję oliwek (zestawienie 4).

ZESTAWIENIE (4) WYNIKÓW IMPLEMENTACJI KOMPUTEROWEJ DLA WSPÓŁCZYNNIKA MARCOTORCHINO

Mierniki statystyczne	Charakterystyka współczynnika Marcotorchino
Wartość współczynnika	0,016
Poziom istotności	0,05
Hipoteza zerowa H_0	Y i Z nie wpływają istotnie na X
p-value	0,004
Wartość statystyki	50,46
Liczba stopni swobody	27

Są podstawy do odrzucenia hipotezy zerowej

Ź r ó d ł o: jak przy tabl. 1.

Przykład 3

Zbadano zadowolenie pacjentów danego oddziału (Z) (zmienna objaśniająca) z lekarzy (X) i pielęgniarek (Y) (zmienne objaśniane). Otrzymane wyniki zestawiono w formie tablicy trójdzielczej $3 \times 3 \times 4$ (tabl. 5).

TABL. 5. TABLICA TRÓJDZIELCZA 3×3×4

Wyszczególnienie	Zadowolenie pacjentów z lekarzy/pielęgniarek oddziałów												razem
	chirurgia			ginekologia			wewnętrzny			ortopedia			
	0	1	2	0	1	2	0	1	2	0	1	2	
0	3	2	0	1	0	1	4	0	1	1	0	0	13
1	6	2	2	4	3	0	4	5	1	7	6	4	44
2	4	9	22	9	20	34	10	16	22	2	8	24	180
Razem	13	13	24	14	23	35	18	21	24	10	14	28	237

U w a g a. 0 — niezadowolony, 1 — prawie zadowolony, 2 — zadowolony.
Ź r ó d ł o: jak przy tabl. 3.

Po utworzeniu tablicy trójdzielczej (tabl. 5) w arkuszu „tablica” (począwszy od komórki B2) uruchomiono procedurę „WspKor3D”. W arkuszu „wyniki” za pomocą pola kombi wybrano opcję „Lombardo” i wyznaczono wartość współczynnika Lombardo, gdy Z jest cechą objaśniającą, a X i Y są cechami objaśnianymi. Następnie założono hipotezę zerową, że oddział (Z) nie wpływa istotnie

na liczbę pacjentów zadowolonych z lekarzy (X) i pielęgniarek (Y). Wprowadzono poziom istotności testu α. Ponieważ $pv_L = 0{,}95 > \alpha$, zatem nie ma podstaw do odrzucenia hipotezy zerowej, czyli oddział szpitalny nie wpływa istotnie na liczbę pacjentów zadowolonych z lekarzy (X) i pielęgniarek (Y) (zestawienie 5).

ZESTAWIENIE (5) WYNIKÓW IMPLEMENTACJI KOMPUTEROWEJ DLA WSPÓŁCZYNNIKA LOMBARDO

Mierniki statystyczne	Charakterystyka współczynnika Lombardo
Wartość współczynnika	0,036
Poziom istotności	0,05
Hipoteza zerowa H_0	Z nie wpływa istotnie na X i Y
p-value	0,949
Wartość statystyki	16,99
Liczba stopni swobody	28

Nie ma podstaw do odrzucenia hipotezy zerowej

Ź r ó d ł o: jak przy tabl. 1.

Podsumowanie

Dysponując gotowym narzędziem do pomiaru siły związku między trzema cechami, które można pobrać z Internetu, czytelnik może samodzielnie przeprowadzić badania na podstawie danych zgromadzonych w formie tablicy trójdzielczej. Pomocne w tym są dokładnie opisane przykłady wraz z ich implementacją komputerową, w której istnieje możliwość wyboru jednego z trzech współczynników siły związku między cechami w tablicy trójdzielczej.

dr Piotr Sulewski — *Akademia Pomorska w Słupsku*

LITERATURA

Beh E., Simonetti B., D'Ambra L. (2007), *Partitioning a non-symmetric measure of association for three-way contingency tables*, „Journal of Multivariate Analysis", Vol. 98 (7)

Cohen J. (1960), *A coefficient of agreement for nominal scales*, „Educational and Psychological Measurement", Vol. 20 (1)

Goodman L., Kruskal W. (1954), *Measures of Association for Cross Classifications*, „Journal of the American Statistical Association", Vol. 49

Gray L. N., Williams J. S. (1975), *Goodman and Kruskal's tau b: multiple and partial analogs*, [in:] *Proceedings of the Social Statistics Section*, American Statistical Association

Light R. J., Margolin B. H. (1971), *An Analysis of Variance for Categorical Data*, „Journal of the American Statistical Association", Vol. 66, No. 335

Lombardo R. (2011), *Three-way association measure decompositions: The Delta index*, „Journal of Statistical Planning and Inference" No. 141

Marcotorchino F. (1984), *Utilisation des Comparaisons par Paires en Statistique des Contingences*, Partie (I), Etude IBM F069, France

Oktaba W. (1974), *Elementy statystyki matematycznej i metodyka doświadczalnictwa*, PWN, Warszawa

Sobczyk M. (1996), *Statystyka*, PWN, Warszawa

Sulewski P. (2007), *Moc tablicy dwudzielczej jako test niezależności*, „Wiadomości Statystyczne", nr 6, GUS

Sulewski P. (2013), *Wielowymiarowe uogólnienie testu niezależności*, „Wiadomości Statystyczne", nr 12, GUS

SUMMARY

In the literature are popular statistical tool to measure the strength of the association between features in the 2-feature tables. However, the mechanisms to measure the strength of the association between the 3 feature in the tables are not too popular, and Polish literature practically does not exist. This paper describes an 3-featur array with its characteristics and provides 3-feature tables theory. The paper presents the computer implementation for measuring the strength of the relationship between the three modes with the Gray-Williams, Marcotorchino and Lombardo factors. The article also presents ready procedures and functions stored in VBA, enabling the reader to carry out independent surveys.

РЕЗЮМЕ

В статистичекой литературе популярными являются инструменты для измерения прочности связи между характристиками в двойных таблицах. Однако механизмы для измерения прочности связи между характристиками в тройных таблицах не очень популярны, а в литературе на польском языке практически не выступают.

В статье была определена тройная таблица вместе с ее характеристикой, а также была представлена теория тройных таблиц. Была представлена компьютерная реализация касающаяся измерения прочности связи между тремя характристиками с использованием коэффициентов: Грэй-Уильямса, Маркоторчино и Ломбардо. Были разработаны готовые процедуры и функции представленные на языке VBA, позволяющие читателю самостоятельно проводить статистические обследования.

Szymon WÓJCIK

Prognozowanie inflacji w Polsce na podstawie modeli autoregresji wektorowej[1]

Prognozowanie wartości inflacji jest kwestią kluczową w prowadzeniu skutecznej polityki pieniężnej. Nie oznacza to jednak, że bank centralny jest jedynym odbiorcą informacji dotyczących przyszłego kształtowania się cen. Prognozy inflacji są wykorzystywane zarówno przez przedsiębiorstwa do opracowywania strategii inwestycyjnej, jak i przez inwestorów na rynkach kapitałowych. Wpływ ten jest szczególnie istotny w konstrukcji portfela inwestycyjnego opartego na konkretnych instrumentach finansowych. Wielkość inflacji ma wpływ zarówno na zyski z obligacji, jak i na koszt zaciągnięcia kredytu. Te zależności są szczególnie związane z wysokością stóp procentowych ustalanych przez bank centralny, co niejako zamyka ten krąg powiązań. Równie istotne wnioski z prognoz inflacji wyciągnąć mogą instytucje związane z obsługą świadczeń emerytalnych. Przy założeniu indeksacji emerytur opartej na stopie inflacji, znajomość przyszłych wartości cen ma bardzo istotne znaczenie w planowaniu budżetu krajowego na kolejne lata.

W przedstawionym badaniu wykorzystano wektorowe modele autoregresyjne (*vector autoregression* — VAR) w celu przetestowania, która z teorii powstawania inflacji jest najlepszym punktem wyjścia do prognozowania zmienności cen. Następnie, w celu sprawdzenia stabilności otrzymanych wyników, przeprowadzono analizę na trzech szeregach czasowych różniących się długością.

W artykule omówiono wcześniejsze badania wykorzystujące metodologię VAR w prognozowaniu inflacji *CPI* dla Polski; przybliżono teorie powstawania inflacji oraz przedstawiono odpowiednie zmienne wykorzystane w modelach; opisano konstrukcję modeli VAR oraz na ich podstawie koncepcję prognozowania.

WCZEŚNIEJSZE BADANIA

Modele VAR, ze względu na swoje dobre własności prognostyczne, są szeroko wykorzystywane w badaniach empirycznych.

Głównym punktem odniesienia opracowania jest publikacja A. Przybylskiej-Mazur (2013), w której autorka stosuje m.in. modele VAR do prognozowania inflacji oraz wskazuje na rolę prognoz w polityce pieniężnej. W opisanym przez autorkę badaniu wykorzystano dane miesięczne wskaźnika inflacji *CPI* z okresu od stycznia 2004 r. do marca 2011 r. Dodatkowo do zbioru zmiennych modelu włączono: stopę referencyjną (wartości z końca miesiąca), WIG (również wartości

[1] Autor pragnie złożyć serdeczne podziękowania drowi Pawłowi Baranowskiemu za naukowe wsparcie przy pisaniu artykułu.

z końca miesiąca), dynamikę produkcji przemysłowej, kurs walutowy (euro) oraz podaż pieniądza *M*3. Okres opracowanych prognoz wynosił od 1 do 5 miesięcy. Wyniki zestawiono z modelami autoregresyjnymi i autokowariancyjnymi dla wskaźnika inflacji oraz z tradycyjnym modelem VAR polityki pieniężnej, zawierającym wskaźnik inflacji, stopę procentową oraz dynamikę produkcji przemysłowej. Spośród wykorzystanych metod najlepsze wyniki otrzymano dla modelu autokowariancyjnego. Autorka podkreśla jednak, że wszystkie modele stanowią skuteczne narzędzie do prognozowania wartości wskaźnika inflacji i mogą być uznane za konkurencyjne względem innych metod prognozowania.

Poza tym o wykorzystywaniu metodologii VAR w prognozowaniu dynamiki cen traktowało badanie P. Baranowskiego, M. Mazurek, M. Nowakowskiego i M. Raczko (2010). Autorzy posłużyli się indeksem inflacji *CPI* oraz jego dwunastoma komponentami, wykorzystując dane o częstotliwości kwartalnej z okresu I 1999 r.—IV 2008 r. Zmienne modelowane były za pomocą modeli VAR oraz jednorównaniowych modeli szeregów czasowych. Choć badanie skoncentrowane było na prognozowaniu szeregów agregatowych i ich składowych, to niektóre jego wyniki przyniosły również istotne przesłanie dotyczące skuteczności modeli VAR. Autorzy oceniają, że w miarę wydłużania okresu prognoz modele VAR zyskują przewagę nad innymi stosowanymi metodami.

Przykład wykorzystania metodologii VAR w charakterze modelu *benchmarkowego* do porównań innych metod prognostycznych podaje artykuł P. Baranowskiego i A. Leszczyńskiej (2011). Modele VAR użyte przez autorów składały się z dwóch równań uwzględniających jako regresory stopę inflacji *CPI* oraz przyrost jednostkowych kosztów pracy. Wykorzystane modele w krótkim okresie generowały równie dobre lub lepsze prognozy w stosunku do reszty modeli tworzonych przez autorów.

ZARYS TEORII I WYKORZYSTANE DANE

W badaniu podjęto próbę spojrzenia na teorie powstawania inflacji. W artykule skupiono się na trzech głównych teoriach powstawania inflacji:
- monetarystycznej,
- popytowej,
- kosztowej.

Teoria monetarystyczna

Współcześnie klasyczną teorię monetarystyczną kojarzy się z noblistą Miltonem Friedmanem, który stwierdził, że: *Inflacja jest zawsze i wszędzie zjawiskiem pieniężnym*[2]. Wynika z tego, że to podaż pieniądza jest głównym czynnikiem

[2] Friedman (1963), s. 29.

decydującym o uruchomieniu i rozwoju inflacji[3]. Teoria monetarystyczna zakłada egzogeniczność podaży pieniądza jako czynnika wywołującego inflację[4]. Oznacza to, że za inflację odpowiedzialna jest polityka monetarna państwa. Naruszenie równowagi monetarnej poprzez zwiększenie podaży pieniądza na rynku skutkuje wzrostem inflacji. Ponadto teoria odrzuca założenie o egzogeniczności szybkości obiegu pieniądza[5]. Monetaryści wykazali, że czynniki mające wpływ na szybkość obiegu pieniądza są względnie stabilne[6]. Istotnym elementem teorii monetarystycznej jest również założenie, że pieniądz może stymulować gospodarkę w krótkim okresie, jednak w długim okresie wzrost produkcji jest zależny wyłącznie od sfery realnej (oszczędności, inwestycji). Zjawisko to zwane jest długookresową neutralnością pieniądza.

Poglądy głoszone przez szkołę monetarystyczną wiodą do stwierdzenia, że tylko nieskrępowany mechanizm wolnorynkowy wiedzie do stabilnego wzrostu gospodarki, a najskuteczniejszym sposobem przeciwdziałania inflacji jest dostosowanie tempa wzrostu podaży pieniądza do wzrostu produkcji realnej.

Teoria popytowa

Popytowa teoria inflacji jest ściśle związana z koncepcją luki inflacyjnej opracowaną przez Johna Maynarda Keynesa. Polega ona na tym, że przy nagłym wzroście jednego ze składników popytu finalnego (np. wydatków państwa) pojawi się popyt niezaspokojony, tzn. nadwyżka popytu nad podażą. Tak więc przy stosunkowo małej podaży będzie silnie działać mechanizm wolnorynkowy, przejawiający się wzmożoną aktywnością nabywców. Mechanizm ten przy braku barier administracyjnych poprowadzi do wzrostu cen. Dalszym efektem tego procesu będzie wzrost dochodów przedsiębiorców, a co za tym idzie spadek realnych płac osób zatrudnionych[7]. W przypadku silnego dążenia grup zawodowych do waloryzacji ich płac realnych będziemy mieli do czynienia z rozwinięciem pojęcia luki inflacyjnej do popytowej teorii inflacji (*demand-pull inflation*). Rezultatem opisanych zjawisk może być pojawienie się spirali inflacyjnej widocznej w zwiększaniu się poziomu cen. Warto podkreślić, że teoria popytowa zakłada całkowite wykorzystanie mocy wytwórczych oraz ewentualne opóźnienia w procesach dostosowawczych.

Teoria kosztowa

Znana jest ona również jako koncepcja inflacji pchanej przez koszty (*cost-push inflation*). Teoria ta podkreśla wpływ rosnących kosztów produkcji na tendencję do wzrostu ogólnego poziomu cen[8]. Podwalin tej teorii doszukiwać się

[3] Kołodko (1987), s. 20.
[4] Pollok (2000), s. 10.
[5] Begg, Fischer, Dornbusch (2007), s. 214.
[6] Kwiatkowski (2005), s. 419.
[7] Welfe (1993), s. 16.
[8] Kołodko (1987), s. 23.

można w doświadczeniach krajów rozwiniętych z lat 50. ub. stulecia, kiedy to inflacja zaczęła sięgać poziomu dwucyfrowego, pomimo niepełnego wykorzystania mocy wytwórczych (pojawienie się bezrobocia). Zjawiska tego nie można było wyjaśnić na podstawie teorii popytowej, toteż zaczęto doszukiwać się źródła tak wysokiej inflacji we wzroście szeroko pojętych kosztów wytworzenia[9]. Zaliczano do nich koszty wynagrodzeń dla pracowników, które rosnąc powodują inflację oraz wzrastające ceny dóbr i surowców (w tym importowanych). Warto wspomnieć, że inflację kosztową generują również różnego rodzaju monopole krajowe lub zagraniczne (np. ceny ropy naftowej)[10], które wykorzystując pozycję monopolisty i chcąc zwiększyć zysk podnoszą ceny wytwarzanych dóbr bez względu na wielkość popytu na nie.

Opisane teorie będą odzwierciedlone w modelach ekonometrycznych przy pomocy trzech zmiennych makroekonomicznych:

- podaż pieniądza *M*3,
- sprzedaż detaliczna towarów według rodzajów działalności przedsiębiorstwa,
- przeciętne wynagrodzenia brutto w sektorze przedsiębiorstw.

Podaż pieniądza M3

Dane na temat pierwszej zmiennej pobrano ze strony internetowej NBP. Zestawienie *Podaż pieniądza M3 i czynniki jego kreacji* podaje w ujęciu miesięcznym podaż agregatu pieniężnego *M*3. Agregat ten zawiera podaż: gotówki w obiegu, depozytów na żądanie, depozytów terminowych oraz dłużnych papierów wartościowych (obligacji) z terminem pierwotnym wykupu do 2 lat. Na podstawie tych danych obliczono wskaźniki dynamiki, przyjmując jako 100 analogiczny okres roku poprzedniego.

Sprzedaż detaliczna towarów według rodzajów działalności przedsiębiorstwa

Dane w cenach stałych pobrano z bazy GUS jako miesięczne wskaźniki z analogicznym okresem roku poprzedniego=100. Odwołując się do definicji GUS można stwierdzić, że zmienna ta zawiera sprzedaż towarów własnych i komisowych (nowych i używanych) w punktach sprzedaży detalicznej, placówkach gastronomicznych oraz innych punktach (np. składy, magazyny) w ilościach wskazujących na zakup dla potrzeb indywidualnych nabywców[11]. Dodatkowo warto podkreślić, że sprzedaż detaliczna przedstawiona jest w cenach realizacji, czyli cenach faktycznie płaconych przez konsumentów, łącznie z podatkiem od towarów i usług (VAT).

[9] Welfe (1993), s. 17.
[10] Pollok (2000), s. 12.
[11] www.stat.gov.pl — zbiór wybranych definicji pojęć społeczno-ekonomicznych z zakresu rynku wewnętrznego.

Średnie wynagrodzenia brutto w sektorze przedsiębiorstw

Informacje zaczerpnięto z publikacji GUS *Zatrudnienie…* (2011) oraz z biuletynów statystycznych. Dane obejmują podmioty działające zarówno w sektorze prywatnym, jak i publicznym. Na podstawie tych danych policzono wskaźniki dynamiki, przyjmując za 100 analogiczny okres roku poprzedniego.

ZESTAWIENIE TEORII INFLACYJNYCH I ODPOWIADAJĄCYCH IM ZMIENNYCH

Teorie inflacyjne	Zmienne
Monetarystyczna	podaż pieniądza *M*3
Popytowa	sprzedaż detaliczna towarów według rodzajów działalności przedsiębiorstwa
Kosztowa	średnie wynagrodzenia brutto w sektorze przedsiębiorstw

Ź r ó d ł o: opracowanie własne.

Wskaźnik cen towarów i usług konsumpcyjnych

Nazywany również indeksem cen konsumenckich, jest najpopularniejszym wyznacznikiem inflacji. Popularność tego wskaźnika wynika głównie z tego, że opisuje on zmiany cen towarów i usług nabywanych przez większość społeczeństwa. Główną ideą jego budowy jest analiza cen ustalonego koszyka dóbr, który jest zarazem systemem wag dla poszczególnych składowych wskaźnika. W statystyce polskiej przyjęto coroczną aktualizację tego koszyka, odzwierciedlającą zmiany struktury spożycia konkretnych dóbr i usług przez konsumentów. Zabieg ten może utrudniać porównywanie ze sobą wskaźników inflacji w poszczególnych latach, ponieważ analizie podlegają wtedy ceny dla różnych koszyków dóbr. Na potrzeby badania dane o tej zmiennej pobrano z baz GUS w postaci miesięcznych indeksów, gdzie analogiczny okres roku poprzedniego=100.

Okres, z którego pochodziły dane wykorzystane w badaniu, to lata 2000—2013, co skutkuje wykorzystaniem 168 miesięcznych obserwacji dla każdej zmiennej. Tak dobrana długość szeregów czasowych pozwala na zapewnienie konkluzywności otrzymanych wyników, jak i na uchwycenie reakcji zmiennych na kryzys w gospodarce światowej z lat 2007—2009.

W dalszych rozważaniach przyjęto następujące oznaczenia:

m — podaż pieniądza *M*3 (ceny bieżące) — dynamika (analogiczny okres roku poprzedniego=100),

s — sprzedaż detaliczna towarów według rodzajów działalności przedsiębiorstwa (w cenach stałych) — dynamika (analogiczny okres roku poprzedniego=100),

w — średnie wynagrodzenia brutto w sektorze przedsiębiorstw (ceny bieżące) — dynamika (analogiczny okres roku poprzedniego=100),

π — wskaźnik cen towarów i usług konsumpcyjnych — dynamika (analogiczny okres roku poprzedniego=100).

W tabl. 1 przedstawiono charakterystykę zmiennych.

TABL. 1. WARTOŚCI MIERNIKÓW STATYSTYCZNYCH WEDŁUG ZMIENNYCH W OKRESIE STYCZEŃ 2000 R.—GRUDZIEŃ 2013 R.

Wyszczególnienie	*m*	*s*	*w*	*π*
Średnia	110,21	105,59	105,50	103,33
Minimum	98,10	82,60	100,00	100,20
Maksimum	120,30	127,70	118,50	111,60
Odchylenie standardowe	4,88	5,97	3,49	2,43
Kurtoza	2,45	4,28	4,35	5,12
Test normalności Jarque'a-Bera	5,17 (*p-value*=0,08[a])	13,97 (*p-value*=0,00)	53,69 (*p-value*=0,00)	88,52 (*p-value*=0,00)

a P-value to graniczny poziom istotności, czyli prawdopodobieństwo popełnienia błędu pierwszego rodzaju.
Ź r ó d ł o: opracowanie własne.

Spośród badanych zmiennych jedynie dynamika podaży pieniądza wykazywała cechy normalności rozkładu na poziomie istotności równym 0,05. Największą zmiennością charakteryzował się szereg dynamiki sprzedaży detalicznej, najmniejszą natomiast — indeks inflacji *CPI*. Wszystkie zmienne oprócz dynamiki cen miały charakter leptokurtyczny, co świadczy o większej koncentracji wokół średniej. Ich rozkład jest bardziej wysmukły w stosunku do rozkładu normalnego. Leptokurtyczność przejawia się również w występowaniu obserwacji bardzo małych lub bardzo dużych. Nieco większe rozproszenie w stosunku do rozkładu normalnego wykazywała tylko dynamika podaży pieniądza.

Opisane zmienne wykorzystano w dalszej analizie do stworzenia modeli ekonometrycznych.

METODOLOGIA BADANIA

Metoda VAR opiera się na trzech zasadniczych przesłankach, tj. na braku[12]:
- podziału na zmienne endo- i egzogeniczne,
- zerowych restrykcji nałożonych na parametry modelu,
- konieczności uwzględniania teorii ekonomicznej przy budowie modelu.

To właśnie ateoretyczność modelu VAR jest jedną z jego najbardziej charakterystycznych cech. Mimo że idea tych modeli była początkowo krytykowana[13], uzyskują one dobre wyniki w:
- prognozowaniu,
- modelowaniu zależności,
- badaniu kointegracji,

[12] Szczegółowy opis konstrukcji VAR podaje Welfe (2009), s. 378—383.
[13] Charemza, Deadman (1997), s. 151—153.

- analizie odpowiedzi na impuls,
- badaniu przyczynowości[14].

Model VAR(p) dla k zmiennych (k równań) można przedstawić wzorem:

$$\boldsymbol{y}_t = \boldsymbol{A}_0 \boldsymbol{d}_t + \sum_{i=1}^{p} \boldsymbol{A}_i \boldsymbol{y}_{t-i} + \boldsymbol{\varepsilon}_t$$

gdzie:

$\boldsymbol{y}_t = [y_{1,t}\ y_{2,t} \ldots y_{k,t}]^T$ — wektor zmiennych endogenicznych,
$\boldsymbol{d}_t$ — wektor zmiennych deterministycznych (w badaniu wektor ten zawierał wyraz wolny),
$\boldsymbol{A}_0$ — macierz parametrów przy zmiennych wektora $\boldsymbol{d}_t$ o wymiarach ($k \times k$),
$\boldsymbol{A}_i$ — macierz parametrów przy i-tym opóźnieniu wektora zmiennych modelu VAR o wymiarach ($k \times k$),
p — rząd opóźnień modelu odzwierciedlający maksymalną długość opóźnień zmiennych endogenicznych,
$\boldsymbol{\varepsilon}_t = [\varepsilon_{1,t}\ \varepsilon_{2,t} \ldots \varepsilon_{k,t}]^T$ — k-wymiarowy wektor (białoszumowych) składników losowych.

Przykładowo, model VAR(1) dla dwóch zmiennych można zapisać jako (wykorzystane oznaczenia stosowane będą w dalszej części artykułu):

$$y_{1,t} = \alpha_{1,0} + \alpha_{1,1}\, y_{1,t-1} + \alpha_{1,2}\, y_{2,t-1} + \varepsilon_{1,t}$$

$$y_{2,t} = \alpha_{2,0} + \alpha_{2,1}\, y_{1,t-1} + \alpha_{2,2}\, y_{2,t-1} + \varepsilon_{2,t}$$

Modele wykorzystywane w badaniu uwzględniały zmienną deterministyczną w postaci wyrazu wolnego. Do estymacji parametrów modelu wektorowej autoregresji wykorzystano klasyczną metodę najmniejszych kwadratów (KMNK)[15] dla każdego równania oddzielnie, ze względu na brak powiązań

[14] Witkowska, Matuszewska, Kompa (2008), s. 140.

[15] Zastosowanie klasycznej metody najmniejszych kwadratów może budzić wątpliwości, w przypadku gdy wykorzystane szeregi czasowe zawierają nietypowe wartości, takie jak na przykład zmiany strukturalne lub okresy kryzysu gospodarczego. Należy jednak zauważyć, że jedną z cech modeli VAR jest bardzo silne dopasowanie do danych, stąd silna reakcja prognoz VAR na nagłe zmiany struktury wykorzystanego szeregu czasowego będzie własnością pożądaną. Formalnie podejmowane są próby opracowywania estymatorów dedykowanych modelom VAR, które wykazują odporność na wartości nietypowe, jednakże w praktyce dominuje podejście oparte na estymatorze KMNK (Jonáš, 2009).

między bieżącymi zmiennymi endogenicznymi. W przypadku występowania powiązań między składnikami losowymi poszczególnych równań stosuje się uogólnioną metodę najmniejszych kwadratów (UMNK)[16]. Co do struktury wykorzystywanych danych należy dodać, że modele VAR w klasycznej postaci wykorzystywane są do analizy zmiennych stacjonarnych. W przypadku gdy mamy do czynienia ze zmiennymi niestacjonarnymi, często stosowane jest przekształcenie modelu VAR w wektorowy model korekty błędem (VECM)[17] lub przekształcenie modelu wyjściowego w model na pierwszych przyrostach.

Z kolei predykcja na podstawie modelu VAR ma charakter mechaniczny. W pierwszym prognozowanym okresie wykorzystywane są oszacowania parametrów modelu oraz obserwacje zmiennych opóźnionych. W kolejnych okresach prognozowanie przebiega sekwencyjnie z wykorzystaniem sformułowanych już prognoz[18].

Schemat badania i dobór opóźnień

W badaniu postanowiono wykorzystać trzy podokresy badawcze. Pierwszy z nich to pełny szereg czasowy — styczeń 2000 r.—grudzień 2013 r. Kolejne to odpowiednio skracane podszeregi — styczeń 2002 r.—grudzień 2013 r. i styczeń 2004 r.—grudzień 2013 r. Takie postępowanie pozwoliło stwierdzić, czy skracanie długości wykorzystanego szeregu czasowego ma znaczący wpływ na wybór najlepszego modelu prognostycznego. Należy dodać, że prognozy poza próbę (*out-of-sample)* przeprowadzono dla okresu weryfikacji 24 miesięcy. Długość szeregu, na podstawie którego estymowano parametry, zwiększała się stopniowo (*recursive sample*, *expanding window*). Na tak przygotowanych szeregach czasowych szacowane były trzy modele główne. Każdy z nich łączył inflację *CPI* z jedną z trzech opisanych zmiennych odpowiadających teoriom powstawania inflacji. Dodatkowo oszacowano cztery modele „mieszane" stanowiące kombinację trzech zmiennych s, m, w i zmiennej π. Są to kolejno modele: $\{\pi, s, m\}$, $\{\pi, m, w\}$, $\{\pi, w, s\}$, $\{\pi, s, m, w\}$. W tabl. 2 zestawiono poszczególne modele w kontekście różnych długości szeregu czasowego i przypisane im opóźnienia modeli VAR. Wyboru opóźnień dokonano na podstawie bayesowskiego kryterium informacyjnego Schwartza (*SBIC*). Kryterium to jest preferowane z uwagi na fakt, że zawiera ono w swojej konstrukcji mechanizm najsilniej „karzący" za ilość szacowanych parametrów modelu. Jest to szczególnie pożądana właściwość w przypadku niewielkiej liczby obserwacji.

[16] Osińska (2006), s. 219.
[17] Witkowska, Matuszewska, Kompa (2008), s. 140.
[18] Kusideł (2000), s. 29 i 30.

TABL. 2. OPÓŹNIENIA WEDŁUG MODELI VAR NA PODSTAWIE *SBIC*

Zmienne w modelach	2000—2013	2002—2013	2004—2013
$\pi\ m$	2	2	1
$\pi\ s$	2	2	2
$\pi\ w$	3	3	3
$\pi\ s\ m$	2	1	1
$\pi\ m\ w$	2	2	1
$\pi\ w\ s$	2	2	1
$\pi\ s\ m\ w$	1	1	1

Ź r ó d ł o: opracowanie własne.

Jak wskazuje tabl. 2, różnica w doborze opóźnień występuje w najkrótszym szeregu czasowym i dotyczy modeli $\{\pi, m\}$, $\{\pi, m, w\}$ oraz $\{\pi, w, s\}$. Model $\{\pi, s, m\}$ jedynie dla najdłuższego szeregu wskazywał na konieczność użycia większej liczby opóźnień.

Aby przybliżyć czytelnikowi strukturę ostatecznych modeli, przedstawiono równania trzech głównych modeli, przyjmując opóźnienia odpowiadające najdłuższemu wykorzystanemu szeregowi czasowemu.

Teoria monetarystyczna

Zmienną dobraną do inflacji *CPI* jest w tym przypadku dynamika podaży pieniądza *M*3. Oto równania modelu VAR(2) dla tej teorii:

$$\pi_t = \alpha_{1,0} + \alpha_{1,1}\pi_{t-1} + \alpha_{1,2}\pi_{t-2} + \alpha_{1,3}m_{t-1} + \alpha_{1,4}m_{t-2} + \varepsilon_{1,t}$$

$$m_t = \alpha_{2,0} + \alpha_{2,1}\pi_{t-1} + \alpha_{2,2}\pi_{t-2} + \alpha_{2,3}m_{t-1} + \alpha_{2,4}m_{t-2} + \varepsilon_{2,t}$$

Teoria popytowa

Zmienną towarzyszącą inflacji *CPI* jest w tym modelu dynamika sprzedaży detalicznej towarów według rodzajów działalności przedsiębiorstwa. Teorię popytową zilustrowano za pomocą modelu VAR(2):

$$\pi_t = \alpha_{1,0} + \alpha_{1,1}\pi_{t-1} + \alpha_{1,2}\pi_{t-2} + \alpha_{1,3}s_{t-1} + \alpha_{1,4}s_{t-2} + \varepsilon_{1,t}$$

$$s_t = \alpha_{2,0} + \alpha_{2,1}\pi_{t-1} + \alpha_{2,2}\pi_{t-2} + \alpha_{2,3}s_{t-1} + \alpha_{2,4}s_{t-2} + \varepsilon_{2,t}$$

Teoria kosztowa

Zmienną wybraną do stworzenia modelu oprócz inflacji *CPI* jest dynamika średnich wynagrodzeń brutto w sektorze przedsiębiorstw. Dla tej teorii opracowano model VAR(3):

$$\pi_t = \alpha_{1,0} + \alpha_{1,1}\,\pi_{t-1} + \alpha_{1,2}\,\pi_{t-2} + \alpha_{1,3}\,\pi_{t-3} + \alpha_{1,4}\,w_{t-1} + \alpha_{1,5}\,w_{t-2} + \alpha_{1,6}\,w_{t-3} + \varepsilon_{1,t}$$

$$w_t = \alpha_{2,0} + \alpha_{2,1}\,\pi_{t-1} + \alpha_{2,2}\,\pi_{t-2} + \alpha_{2,3}\,\pi_{t-3} + \alpha_{2,4}\,w_{t-1} + \alpha_{2,5}\,w_{t-2} + \alpha_{2,6}\,w_{t-3} + \varepsilon_{2,t}$$

WYNIKI

Prognozy tworzone były dla trzech okresów wynoszących odpowiednio 1, 3 i 6 miesięcy. Weryfikację dokładności prognoz przeprowadzono na podstawie błędów prognoz *ex post — ME*, *MAE* i *RMSE*[19].

Okres miesięczny

Wyniki błędów prognoz dla okresu miesięcznego zestawiono w tabl. 3.

TABL. 3. OSZACOWANIA BŁĘDÓW PROGNOZ *EX POST* DLA JEDNOMIESIĘCZNEGO OKRESU PROGNOZY

Zmienne w modelach	2000—2013			2002—2013			2004—2013		
	ME	*MAE*	*RMSE*	*ME*	*MAE*	*RMSE*	*ME*	*MAE*	*RMSE*
π m	0,08	0,32	0,38	0,09	0,33	0,39	0,18	0,34	0,41
π s	0,03	0,26	0,36	0,03	0,27	0,36	0,05	0,28	0,37
π w	0,04	0,29	0,37	0,06	0,30	0,38	0,08	0,31	0,39
π s m	0,04	0,24	0,33	0,06	0,26	0,34	0,11	0,28	0,36
π m w	0,04	0,30	0,38	0,05	0,31	0,39	0,15	0,33	0,40
π w s	0,03	0,27	0,36	0,03	0,27	0,37	0,08	0,27	0,35
π s m w	0,04	0,24	0,34	0,06	0,26	0,34	0,11	0,29	0,37

Ź r ó d ł o: jak przy tabl. 1.

Rozpatrując trzy główne modele (odpowiadające teoriom powstawania inflacji) można stwierdzić, że najlepsze własności prognostyczne miał model teorii

[19] *ME* — średni błąd (*mean error*) definiowany jako średnia arytmetyczna odchyleń wartości oszacowanych od wartości rzeczywistych; *MAE* — tzw. średni błąd absolutny (*mean absolute error*) różni się on od poprzednika tym, że jest średnią z wartości bezwzględnych różnic między wartościami oszacowanymi i rzeczywistymi; *RMSE* — pierwiastek średniego kwadratu błędu (*root mean square error*) powstaje poprzez podniesienie do kwadratu różnic i wyciągnięcie pierwiastka kwadratowego z otrzymanej średniej. Szczegóły konstrukcji tych błędów oraz inne miary dokładności prognoz podaje Welfe (2009), s. 245—251.

popytowej (model π *s*), najgorzej zaś prognozował model monetarystyczny uwzględniający inflację *CPI* oraz dynamikę podaży pieniądza *M*3 (model π *m*). Wynik ten zaobserwować można w przypadku wszystkich trzech szeregów czasowych.

Uwzględnienie w porównaniu jakości prognoz dodatkowych czterech modeli „mieszanych" pozwoliło stwierdzić, że w przypadku okresu jednomiesięcznego najlepsze wyniki uzyskano na podstawie modelu uwzględniającego inflację *CPI*, dynamikę podaży pieniądza *M*3 oraz dynamikę sprzedaży detalicznej (model π *s m*). Poza tym wyniki poszczególnych modeli nie różniły się znacząco. Na uwagę zasługuje fakt, że wraz ze skracaniem szeregu czasowego wyniki prognoz ulegały pogorszeniu. Jest to zgodne z intuicją badawczą sugerującą wykorzystywanie jak najdłuższych szeregów obserwacji. Dodatnie wartości błędu *ME* dla wszystkich szeregów czasowych wskazują na tendencję modeli do przeszacowywania prognoz (dodatnie obciążenie). Duża różnica pomiędzy wartościami błędów *ME* i *MAE* świadczy o różnokierunkowym charakterze błędów. Znaczy to, że niektóre prognozowane wartości były niższe od obserwacji empirycznych, a niektóre wyższe. Zależności te widoczne były we wszystkich okresach prognozy.

Okres trzymiesięczny

Dla okresu trzymiesięcznego oszacowane błędy prognoz przedstawia tabl. 4.

TABL. 4. OSZACOWANIA BŁĘDÓW PROGNOZ *EX POST* DLA TRZYMIESIĘCZNEGO OKRESU PROGNOZY

Zmienne w modelach	2000—2013			2002—2013			2004—2013		
	ME	*MAE*	*RMSE*	*ME*	*MAE*	*RMSE*	*ME*	*MAE*	*RMSE*
π *m*	0,36	0,61	0,69	0,34	0,61	0,70	0,51	0,67	0,83
π *s*	0,19	0,54	0,60	0,17	0,54	0,60	0,26	0,58	0,63
π *w*	0,14	0,50	0,58	0,20	0,54	0,62	0,31	0,57	0,67
π *s m*	0,22	0,53	0,61	0,28	0,56	0,64	0,40	0,63	0,74
π *m w*	0,20	0,53	0,63	0,23	0,56	0,66	0,43	0,63	0,80
π *w s*	0,19	0,54	0,60	0,18	0,54	0,61	0,33	0,60	0,68
π *s m w*	0,21	0,52	0,60	0,26	0,55	0,63	0,38	0,63	0,74

Ź r ó d ł o: jak przy tabl. 1.

W przypadku tego okresu, rozpatrując trzy główne modele prognostyczne stwierdzono, że również model zawierający stopę inflacji oraz sprzedaż detaliczną (model π *s*) prognozował najlepiej — z wyjątkiem przypadku najdłuższego szeregu czasowego, gdzie najlepszy wynik osiągnął model teorii kosztowej (model π *w*). Uwzględniając cztery modele „mieszane" trudno wyciągnąć jednoznaczne wnioski co do przewagi jednego wybranego modelu nad innymi, z uwagi na niewielkie różnice w wielkości błędów pomiędzy poszczególnymi

modelami. Jeżeli chodzi o model najgorzej prognozujący wskaźnik inflacji *CPI*, to jest to w okresie trzymiesięcznym model dedykowany monetarystycznej teorii powstawania inflacji (model $\pi\ m$). W trzymiesięcznym okresie prognozy również widoczny jest spadek dokładności prognoz wraz ze skracaniem szeregu czasowego wykorzystanego w badaniu. W stosunku do prognoz jednomiesięcznych zaobserwowano ogólny spadek dokładności przewidywań.

Okres sześciomiesięczny

Błędy oszacowane dla sześciomiesięcznego horyzontu prognozy przedstawia tabl. 5.

TABL. 5. OSZACOWANIA BŁĘDÓW PROGNOZ *EX POST* DLA SZEŚCIOMIESIĘCZNEGO OKRESU PROGNOZY

Zmienne w modelach	2000—2013			2002—2013			2004—2013		
	ME	*MAE*	*RMSE*	*ME*	*MAE*	*RMSE*	*ME*	*MAE*	*RMSE*
$\pi\ m$	0,57	0,81	1,11	0,49	0,88	1,11	0,86	0,92	1,34
$\pi\ s$	0,37	0,70	0,93	0,29	0,70	0,91	0,47	0,73	1,00
$\pi\ w$	0,14	0,69	0,84	0,23	0,74	0,92	0,43	0,81	1,01
$\pi\ s\ m$	0,45	0,72	0,99	0,53	0,72	1,04	0,75	0,82	1,22
$\pi\ m\ w$	0,29	0,72	0,92	0,29	0,80	0,99	0,75	0,86	1,24
$\pi\ w\ s$	0,30	0,67	0,89	0,25	0,68	0,90	0,62	0,76	1,10
$\pi\ s\ m\ w$	0,40	0,69	0,95	0,48	0,70	1,01	0,71	0,80	1,18

Ź r ó d ł o: jak przy tabl. 1.

Biorąc pod uwagę jedynie trzy główne modele prognostyczne, w przypadku okresu sześciomiesięcznego nieznaczną przewagę uzyskał model popytowy (model $\pi\ s$). Przewaga ta była widoczna wszakże jedynie w dwóch krótszych podszeregach. Dla najdłuższego szeregu czasowego jednoznaczną przewagę wykazywał model kosztowy (model $\pi\ w$). Włączając do analizy modele „mieszane" nie stwierdzono jednoznacznej przewagi żadnego z dołączonych modeli nad modelem kosztowym lub popytowym. W większości przypadków również obserwowano efekt pogarszania się dokładności prognoz w miarę zmniejszania się długości wykorzystanego szeregu czasowego. Najmniej dokładny z modeli ponownie okazał się model odpowiadający monetarystycznej teorii powstawania inflacji (model $\pi\ m$).

Podsumowanie

Celem badania jest sprawdzenie możliwości prognozowania inflacji *CPI* w Polsce z wykorzystaniem modeli VAR z różnymi zestawami zmiennych, na jeden, trzy i sześć miesięcy. Zmienne te były dobrane tak, aby odpowiadały konkretnym teoriom powstawania inflacji. Ponadto zbadano odporność otrzymanych wyników na zmianę długości wykorzystywanego szeregu czasowego.

Wyniki otrzymane na podstawie trzech głównych modeli wskazały na najlepsze własności prognostyczne modelu teorii popytowej (model uwzględniający stopę inflacji oraz dynamikę sprzedaży detalicznej). Jedynie w przypadku uwzględnienia najdłuższego szeregu czasowego w okresie sześciomiesięcznym model popytowy ustąpił modelowi kosztowemu (model zawierający stopę inflacji oraz dynamikę średnich wynagrodzeń w sektorze przedsiębiorstw). Jednoznacznie najgorsze wyniki prognostyczne otrzymano na podstawie modelu monetarystycznego (zawierającego stopę inflacji oraz dynamikę podaży pieniądza *M*3).

Badanie pokazało, że wraz ze skracaniem długości szeregu czasowego danych modele VAR traciły dokładność (zwiększały się błędy prognoz *ex post*). Ponadto wraz z wydłużaniem okresu prognozy dokładność wyników spadała. W znacznej większości przypadków zwiększenie okresu prognozy, jak i zmiana długości wykorzystanego szeregu czasowego nie miały wpływu na ostateczne wnioski co do przewagi jednych modeli nad innymi. Świadczy to o stabilności otrzymanych wyników.

Warto wspomnieć, że kwestię podobną do opisywanej (krótkoterminowe prognozy inflacji dla Polski z wykorzystaniem różnych zestawów zmiennych) rozważał m.in. G. Szafrański (2011). Pośród wielu zmiennych wykorzystał on również podaż pieniądza do konstrukcji swoich modeli. Pomijając różnice w zastosowanych metodologiach, jego wyniki wykazały duże wartości błędów dla zmiennej agregatowej uwzględniającej podaż pieniądza, co potwierdzają wyniki otrzymane w badaniu. Niemniej jednak na tej podstawie nie możemy wyciągać daleko idących wniosków na temat zasadności stosowania zmiennej „podaż pieniądza" w prognozowaniu inflacji, ponieważ zakres badań oraz wykorzystywane metody i modele ekonometryczne są diametralnie różne.

mgr Szymon Wójcik — *Uniwersytet Łódzki*

LITERATURA

Baranowski P., Leszczyńska A. (2011), *Prognozowanie inflacji w oparciu o hybrydową krzywą Phillipsa dla gospodarki zamkniętej i małej gospodarki otwartej*, „Materiały i Studia", wyd. NBP, Warszawa

Baranowski P., Mazurek M., Nowakowski M., Raczko M. (2010), *Czy dezagregacja indeksu cen poprawia prognozy polskiej inflacji?*, „Przegląd Statystyczny", Zeszyt 1/2010, t. 57, Toruń

Begg D., Fischer S., Dornbusch R. (2007), *Makroekonomia*, Polskie Wydawnictwo Ekonomiczne, Warszawa

Charemza W., Deadman D. (1997), *Nowa ekonometria*, PWE, Warszawa

Friedman M. (1963), *Inflation: Causes and Consequences*, „Dollars and Deficits", Prentice-Hall

Jonáš P. (2009), *Robust Estimation of the VAR Model*, „WDS'09 Proceedings of Contributed Papers", Part I

Kołodko G. W. (1987), *Polska w świecie inflacji*, Książka i Wiedza, Warszawa

Kusideł E. (2000), *Modele wektorowo-autoregresyjne VAR — metodologia i zastosowania*, Absolwent, Łódź

Kwiatkowski E. (2005), *Inflacja*, [w:] Milewski R., Kwiatkowski E. (red.), *Podstawy ekonomii*, Polskie Wydawnictwo Naukowe, Warszawa

Osińska M. (2006), *Ekonometria finansowa*, PWE, Warszawa

Pollok A. (2000), *Inflacja w teorii ekonomii*, Akademia Ekonomiczna w Krakowie

Przybylska-Mazur A. (2013), *Metody prognozowania inflacji a decyzje polityki pieniężnej*, Uniwersytet Ekonomiczny w Katowicach

Szafrański G. (2011), *Krótkoterminowe prognozy polskiej inflacji w oparciu o wskaźniki wyprzedzające*, „Materiały i Studia", nr 263, Warszawa

Welfe A. (1993), *Inflacja i rynek*, PWE, Warszawa

Welfe A. (2009), *Ekonometria. Metody i ich zastosowanie*, PWE, Warszawa

Witkowska D., Matuszewska A., Kompa K. (2008), *Wprowadzenie do ekonometrii dynamicznej i finansowej*, SGGW, Warszawa

Zatrudnienie i wynagrodzenia w gospodarce narodowej w I—III kwartale 2011 r. (2011), GUS

SUMMARY

The article presents a usage of vector autoregressive models in forecasting polish consumer price index. Macroeconomic variables used in this paper are considered to reflect particular economic theories describing causes of inflation. Out-of-sample methodology was used in forecasting process. Accuracy of results was diagnosed by using ex post forecasting errors.

РЕЗЮМЕ

В статье были использованы модели векторной авторегрессии для прогнозирования месячного показателя потребительских цен в Польше. Выбор используемых макроэкономических переменных соответствовал трем теориям формирования инфляции: монетаристской, кейнсианской (курсовой) и издержек. В прогнозировании была использована концепция вне выборки (out-of-sample), а качество результатов было обследовано с использованием ошибок прогноз ex post.

Bartosz TOTLEBEN

Empiryczna analiza wpływu wybranych czynników na wzrost gospodarczy

Problem wzrostu gospodarczego i jego przyczyny są przedmiotem częstych badań ekonomistów. Obecnie dostępne są coraz dłuższe szeregi czasowe danych, z tego względu analizy te powinny być poddawane ciągłej aktualizacji. W pracy zaprezentowano wybrane determinanty wzrostu gospodarczego, przeprowadzono badanie wpływu wybranych zmiennych na PKB *per capita*, dokonano też krótkiego przeglądu literatury tematu. Estymacja modeli różniących się przyjętą metodą badawczą pozwala na wyciągnięcie wniosków z zebranych danych.

Celem opracowania jest określenie wpływu: wydatków rządowych, inflacji, skolaryzacji, bezpośrednich inwestycji zagranicznych, innowacji oraz wolności osobistej i politycznej na PKB *per capita*. Postawiono hipotezę, że wymienione zmienne są determinantami osiąganego poziomu PKB *per capita*.

W badaniu wykorzystano regresję liniową (zastosowano klasyczną metodę najmniejszych kwadratów — KMNK) oraz test przyczynowości Grangera. Poddano analizie dane zebrane z 59 krajów w okresie 1961—2011[1].

Inspiracją do przeprowadzenia badań były publikacje, które opracowali Barro (1991) i Sala-i-Martin (2013), na podstawie tych lektur dokonano doboru zmiennych.

WYBÓR ZMIENNYCH I METODOLOGIA

W badaniu, jak już wspomniano, wzięto pod uwagę wybrane determinanty wzrostu gospodarczego: wydatki rządowe, inflację, skolaryzację, bezpośrednie inwestycje zagraniczne (BIZ), liczbę zgłoszonych patentów oraz wskaźniki wolności FreedomHouse. Źródłem tych danych były bazy *World Development Indicators* (pochodząca ze zbiorów Banku Światowego) oraz *Freedom in the World* (udostępniana przez organizację FreedomHouse).

Poddano analizie 59 krajów, które podzielono zgodnie z klasyfikacją Banku Światowego na 4 statusy dochodowe (tabl. 1). Aby ograniczyć wpływ krótkotrwałych odchyleń do obliczeń wykorzystano średnie 3-letnie[2], co skutkowało

[1] Ze względu na brak danych dla zmiennych niezależnych „skolaryzacja i bezpośrednie inwestycje zagraniczne" przyjęto lata 1970—2011 oraz dla instytucji państwa lata 1973—2011.

[2] W przypadku braku jednego rekordu obliczano średnią z dwóch pozostałych; jeśli brakowało co najmniej dwóch rekordów pozostawiono rekord pusty.

utworzeniem 17 szeregów czasowych[3]. Jedynym kryterium doboru krajów była dostępność danych w wymienionych bazach danych.

TABL. 1. PODZIAŁ KRAJÓW WEDŁUG GRUP DOCHODOWYCH

Grupy dochodowe	Kraje	PKB *per capita* w 2012 r. w USD
1 — niski dochód	Bangladesz, Benin, Burkina Faso, Burundi, Czad, Demokratyczna Republika Kongo (dawny Zair), Gambia, Kenia, Madagaskar, Niger, Ruanda, Togo, Uganda, Zimbabwe	≤1035
2 — średnioniski dochód	Boliwia, Egipt, Filipiny, Gwatemala, Honduras, Indie, Indonezja, Lesoto, Maroko, Pakistan, Paragwaj, Salwador, Sri Lanka, Syria, Zambia	1036—4085
3 — średniowysoki dochód	Algieria, Argentyna, Brazylia, Ekwador, Iran, Kolumbia, Kostaryka, Malezja, Meksyk, Panama, Peru, Tajlandia, Tunezja, Turcja, Węgry	4086—12615
4 — wysoki dochód	Austria, Dania, Finlandia, Francja, Grecja, Hiszpania, Holandia, Irlandia, Japonia, Kanada, Korea Południowa, Norwegia, Portugalia, Szwecja, Urugwaj	≥12616

Ź r ó d ł o: klasyfikacja Banku Światowego.

W opracowaniu przytoczono definicje zmiennych, sposób ich obliczania, zmienność w czasie, a także współczynniki regresji liniowej:

$$y_t = \alpha x_{1_{t-1}} + c \tag{1}$$

gdzie:

y_t — PKB *per capita* w okresie t,
$x_{1_{t-1}}$ — wybrana determinanta w okresie t–1,
c — stała.

Wszystkie równania regresji poddano estymowaniu KMNK, za poziom ufności przyjęto $u = 95\%$.

Zgodnie z definicją proces X jest przyczyną (w sensie Grangera) procesu Y, jeśli dokładniej można przewidywać wielkość Y z wykorzystaniem X niż bez niego (Granger, 1969). W praktyce oznacza to wykorzystanie opóźnionego X pod warunkiem jego istotności statystycznej.

W badaniu postawiono hipotezę zerową, że żadna z opisanych wcześniej zmiennych (wydatki rządowe, inflacja, współczynnik skolaryzacji[4], BIZ, liczba

[3] Przyjmując: t_1 — lata 1961—1963, t_2 — 1964—1966, …, t_{17} — 2009—2011.

[4] Będący arytmetyczną średnią współczynników skolaryzacji dla wszystkich trzech etapów nauczania — podstawowego, średniego i wyższego.

złożonych patentów, wskaźnik wolności FreedomHouse) nie jest przyczyną w sensie Grangera wielkości PKB *per capita*. Aby sfalsyfikować hipotezę, posłużono się następującymi równaniami:

$$y_{i,t} = A_0 D_t + \sum_{j=1}^{q} \gamma_j y_{i,t-j} + \alpha_i + \varepsilon_{i,t} \tag{2}$$

$$y_{i,t} = A_0 D_t + \sum_{j=1}^{q} \gamma_j y_{i,t-j} + \sum_{j=1}^{q} \beta_j x_{k_{i,t-j}} + \alpha_i + \varepsilon_{i,t} \tag{3}$$

gdzie:

$y_{i,t}$ — PKB *per capita* w kraju i (w roku t),
$A_0 D_t$ — deterministyczna część modelu,
x_k — badana zmienna niezależna,
α_i — efekt stały charakterystyczny dla kraju,
ε — błąd losowy.

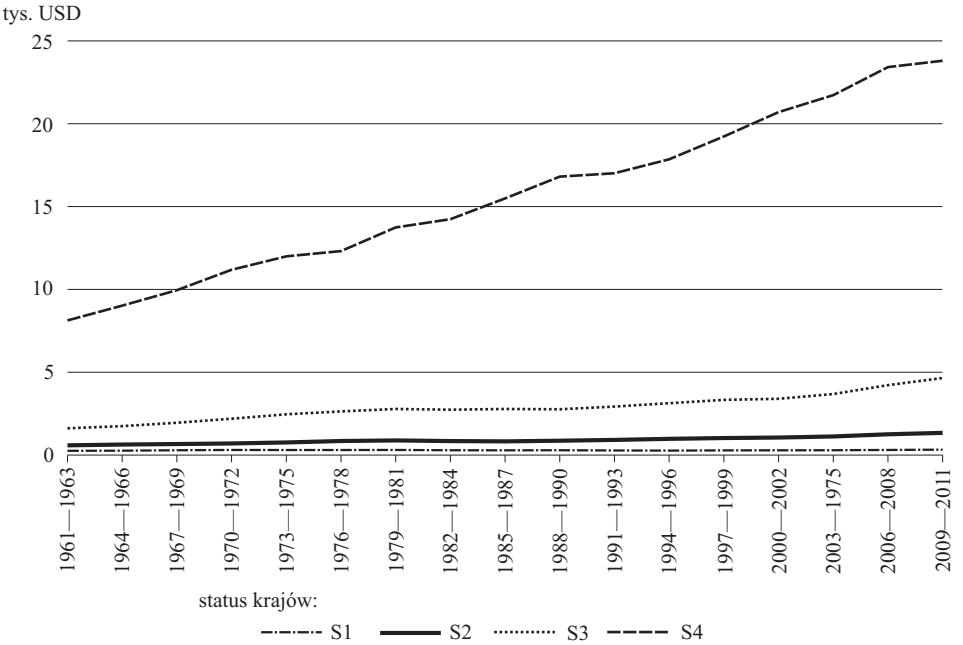
Równanie (2) wykorzystano do zmierzenia, na ile zmienna zależna (PKB *per capita*) może być objaśniona jedynie opóźnionymi zmiennymi zależnymi. Równanie (3) stanowi rozszerzenie równania (2), które wyprowadzono w celu przetestowania, jak napływ nowych informacji (opóźnionej zmiennej niezależnej) oddziałuje na predykcję zmiennej zależnej. Hipotezę zerową można zatem zapisać: $\beta_1 = 0, \ldots, \beta_q = 0$, co oznacza, że przyczynowość w sensie Grangera nie występuje, jeśli równanie (2) pozwala wnioskować o wysokości PKB *per capita* oraz równanie (3) charakteryzuje się brakiem istotności statystycznej co najmniej jednej zmiennej niezależnej.

Do estymacji wybrano uogólnioną metodę momentów (Blundell, Bond, 1998) oraz dwustopniową estymację z korekcją Windmeijera (2005). Wszystkie wartości zmiennych przekształcono logarytmem naturalnym. Otrzymane równanie (3) poddano testowi przyczynowości w sensie Grangera za pomocą testu Walda, za poziom ufności przyjęto 95% oraz użyto statystyki chi-kwadrat. Podobny sposób przeprowadzania testu przyczynowości w sensie Grangera przedstawili Piątek, Pilc i Szarzec (2013). Testując estymowane równania posłużono się maksymalnie opóźnieniami o 3 okresy.

Na wykr. 1 zaprezentowano zmiany w dochodzie przypadającym na osobę w latach 1961—2011 z podziałem na wymienione kategorie dochodowe.

Największym względnym wzrostem PKB *per capita* charakteryzowały się kraje o wysokim dochodzie (wzrost o 193%), średniowysokim (188%) oraz średnioniskim (127%). Najmniej dynamiczny wzrost wystąpił w krajach o niskim dochodzie (zaledwie 26%).

Wykr. 1. PKB *PER CAPITA*



U w a g a. W tysiącach dolarów amerykańskich w cenach stałych z 2000 r.
Ź r ó d ł o: opracowanie własne na podstawie danych Banku Światowego.

DETERMINANTY WZROSTU GOSPODARCZEGO

Wydatki rządowe

Zmienna wydatki rządowe zawiera wszystkie wydatki instytucji rządowych i samorządowych na spożycie ostateczne. Obejmuje wydatki na zakup towarów i usług, wynagrodzenia pracowników oraz większość wydatków na obronę narodową i bezpieczeństwo, nie obejmuje natomiast wydatków przeznaczonych na cele wojskowe. Zmienną tę przedstawiono w relacji do PKB.

Na przestrzeni badanych lat najwyższe wartości wydatków rządowych notowano w krajach o najwyższych dochodach, najniższe zaś w krajach charakteryzujących się najniższym PKB *per capita*. Wyjątek stanowiły lata 1976—1993, gdy obserwowano najniższy udział konsumpcji rządowej w PKB w krajach o średniowysokim dochodzie.

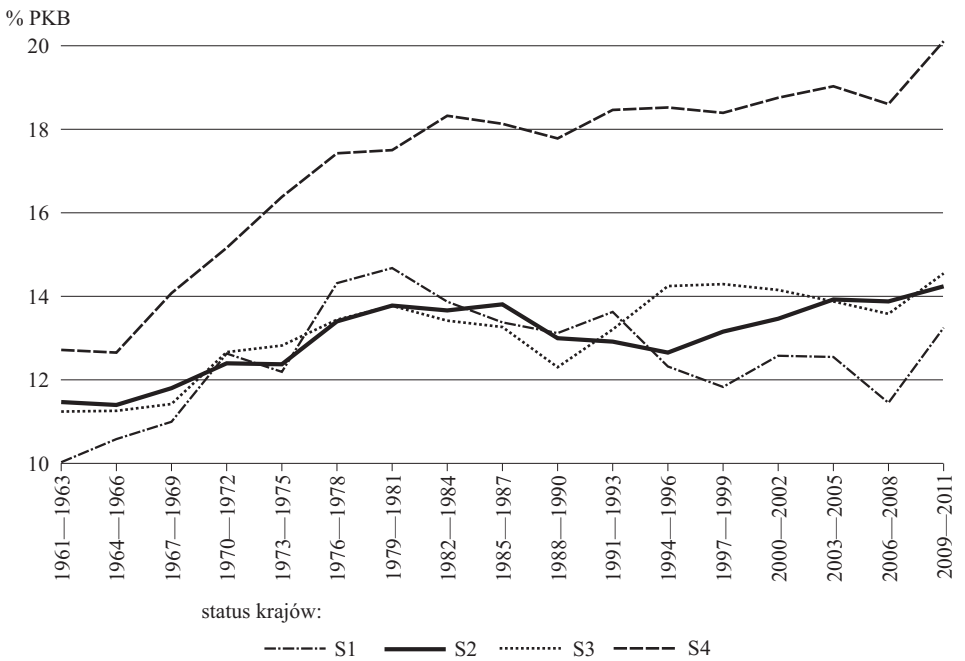
W celu określenia kierunku wpływu wydatków rządowych na PKB *per capita* utworzono cztery modele regresji liniowej. Zmienną niezależną opóźniono o jeden okres. Wielkość opóźnień we wszystkich modelach przyjęto *a priori*. Wyniki zaprezentowano w tabl. 2.

TABL. 2. WSPÓŁCZYNNIKI REGRESJI LINIOWEJ ZMIENNEJ „WYDATKI RZĄDOWE"

Status krajów	Współczynnik zmiennej niezależnej (t–1), (p-*value*)	Współczynnik zmiennej stałej	R^2 w %
*S*1	0,38[a] (0,906)	291,52	0,10
*S*2	166,93 (0,002)	–1238,56	50,38
*S*3	562,83 (0,000)	–4390,01	56,61
*S*4	1874,97 (0,000)	–15722,70	73,82

a Zmienna nieistotna statystycznie.
Ź r ó d ł o: opracowanie własne na podstawie danych Banku Światowego.

Wykr. 2. WYDATKI RZĄDOWE



Ź r ó d ł o: jak przy wykr. 1.

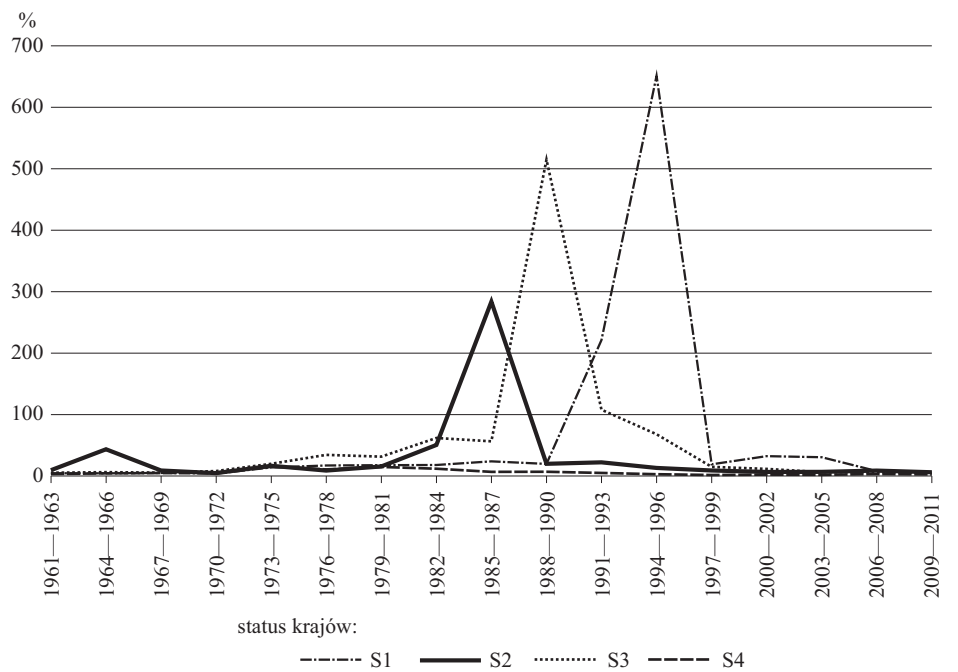
Dodatnie współczynniki opóźnionej zmiennej niezależnej potwierdzają, że wydatki rządowe wpływają pozytywnie na PKB *per capita* w krajach o wysokim, średniowysokim oraz średnioniskim dochodzie. Jedynie dla grupy krajów o najniższych dochodach, ze względu na brak istotności współczynnika x_1, nie było możliwe oznaczenie wpływu.

W literaturze przedmiotu nie występuje zgodność dotycząca badanego problemu. Przytaczając niektóre opinie, Pan (2013) twierdzi, że wzrost konsumpcji rządowej przynosi efekty krótkookresowe w postaci wyższego PKB, jednak niemożliwe jest potwierdzenie tej zależności w długim okresie. Al. Bataineh (2012), wykorzystując modele regresji liniowej, opisał pozytywny wpływ wydatków rządowych na wzrost gospodarczy (na przykładzie Jordanii), co jest zgodne z otrzymanymi wynikami. Z kolei Kormendi i Meguire (1985) nie znaleźli, na podstawie danych empirycznych, dowodów na wzajemną zależność między wydatkami rządowymi a wzrostem gospodarczym. Zarówno Barro (1990), jak i Landau (1983) wskazują na negatywny wpływ wzrostu konsumpcji rządowej na stopę realnego wzrostu PKB *per capita*, ale nie przesądzają o braku pozytywnego wpływu na ogólny dobrobyt.

Inflacja

Wskaźnik inflacji — *CPI* (*Consumer Price Index*) odzwierciedla zmiany kosztów życia przeciętnego konsumenta. Koszyk zawiera średnią ważoną zmian cen najczęściej nabywanych dóbr i usług zgodnie ze wzorem Laspeyere'a.

Wykr. 3. INFLACJA (rok do roku)



Ź r ó d ł o: jak przy wykr. 1.

W analizowanym okresie najniższymi wartościami i najmniejszą zmiennością (liczoną wskaźnikiem zmienności[5]) charakteryzowały się kraje o najwyższym

[5] To jest: $V(S1) \approx 2{,}47$; $V(S2) \approx 2{,}11$; $V(S3) \approx 2{,}14$; $V(S4) \approx 0{,}71$.

dochodzie. W pozostałych grupach dochodowych wyższe współczynniki zmienności odzwierciedlały kryzys inflacyjny, który dotyczy m.in.: Demokratycznej Republiki Konga (1991—2001), Boliwii (1982—1985), Peru (1988—1991), Argentyny (1975—1991) oraz Brazylii (1981—1994).

TABL. 3. WSPÓŁCZYNNIKI REGRESJI LINIOWEJ ZMIENNEJ „INFLACJA"

Status krajów	Współczynnik zmiennej niezależnej (t–1), (*p-value*)	Współczynnik zmiennej stałej	R^2 w %
*S*1	–0,03[a] (0,273)	298,27	8,50
*S*2	–0,40[a] (0,614)	936,30	1,86
*S*3	–0,03[a] (0,989)	2966,64	0,00
*S*4	–441,08[a] (0,111)	18960,65	17,13

a Zmienna nieistotna statystycznie.
Ź r ó d ł o: jak przy tabl. 2.

Bez względu na grupę dochodową otrzymano ujemne wskaźniki zmiennej niezależnej. Jednak należy zaznaczyć, że wszystkie otrzymane równania regresji mają zmienną nieistotną statystycznie, co uniemożliwia prawidłowe wnioskowanie.

Na jednoznacznie negatywny wpływ hiperinflacji uwagę zwracają Bruno i Easterly (1995), choć uznają, że niska i kontrolowana inflacja może (w zależności od innych warunków) być przyczyną lub ograniczeniem wzrostu gospodarczego. Na szkodliwy wpływ inflacji na wzrost gospodarczy zwrócił uwagę m.in. Barro (1991). Pozytywną i wzajemną zależność pomiędzy wzrostem gospodarczym i inflacją opisali Jaradat (2013) oraz Chowdhury i Mallik (2001), którzy zwrócili jednocześnie uwagę, że hiperinflacja uniemożliwia osiągnięcie wzrostu.

Skolaryzacja

Współczynnik skolaryzacji jest to stosunek liczby osób zarejestrowanych jako uczące się na danym poziomie (bez względu na wiek) do liczebności populacji grupy wiekowej powiązanej oficjalnie z owym poziomem edukacji. W opracowaniu przyjęto następujące oznaczenia wskaźnika skolaryzacji edukacji: podstawowy — *EDU*1 (x_1), średni — *EDU*2 (x_2) i wyższy — *EDU*3 (x_3).

Celem zbadania wpływu skolaryzacji na wzrost gospodarczy estymowano następujące typy równań:

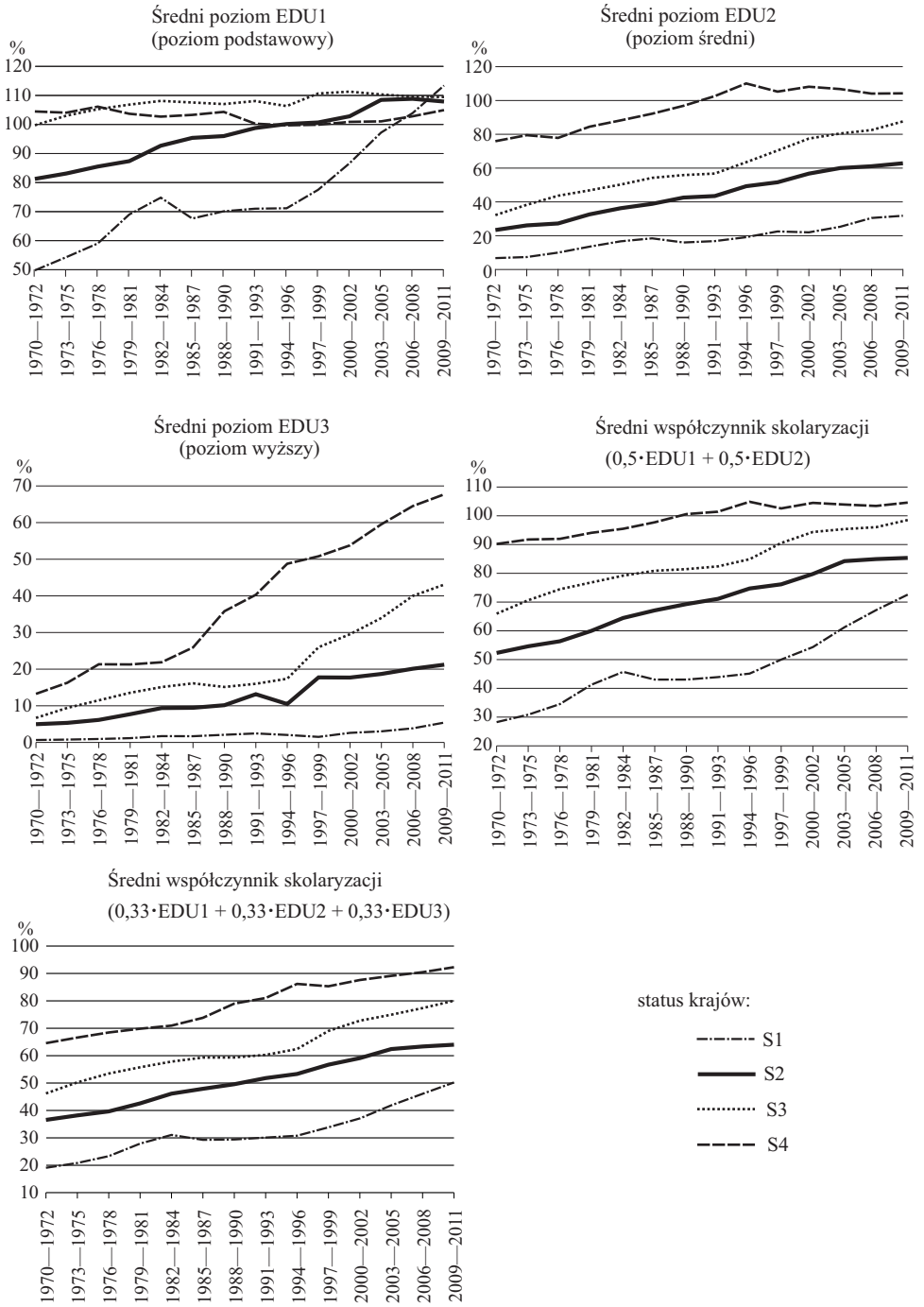
$$y_t = x_{n_{t-1}} + c \quad \text{dla} \quad n \in (1, 2, 3) \tag{4}$$

$$y_t = \left(\overline{x_{1_{t-1}} + x_{2_{t-1}}}\right) + c \tag{5}$$

$$y_t = \left(\overline{x_{1_{t-1}} + x_{2_{t-1}} + x_{3_{t-1}}}\right) + c \tag{6}$$

Opisane współczynniki przedstawiono na wykr. 4.

Wykr. 4. WSKAŹNIKI SKOLARYZACJI



Ź r ó d ł o: jak przy wykr. 1.

Generalnie najwyższym poziomem skolaryzacji (bez względu na przyjętą miarę) charakteryzują się kraje o najwyższym dochodzie, gdzie wartości wskaźników skolaryzacji edukacji podstawowej i średniej oscylują ok. 100%, zaś w przypadku edukacji wyższej ponad 60%. W krajach o najmniejszym PKB *per capita* jedynie dostęp do edukacji podstawowej jest na poziomie ok. 100%, z edukacji średniej w latach 2009—2011 korzystało przeciętnie 31% odpowiedniej grupy wiekowej, zaś z edukacji wyższej — 5,43%.

TABL. 4. WSPÓŁCZYNNIKI REGRESJI LINIOWEJ WSKAŹNIKÓW SKOLARYZACJI PODSTAWOWEJ, ŚREDNIEJ I WYŻSZEJ

Status krajów	Typ równania	Współczynnik zmiennej niezależnej (t–1), (*p-value*)	Współczynnik zmiennej stałej	R^2 w %
*S*1	(4): $n = 1$	0,24[a] (0,411)	280,14	6,22
	$n = 2$	0,27[a] (0,686)	293,10	1,54
	$n = 3$	2,30[a] (0,640)	293,38	2,06
	(5)	0,28[a] (0,485)	284,77	4,53
	(6)	0,41[a] (0,490)	285,12	4,44
*S*2	(4): $n = 1$	17,38 (0,000)	–678,17	81,87
	$n = 2$	12,59 (0,000)	453,89	85,52
	$n = 3$	30,14 (0,000)	630,67	84,92
	(5)	14,63 (0,000)	–26,18	84,58
	(6)	17,92 (0,000)	89,16	85,89
*S*3	(4): $n = 1$	141,57 (0,008)	–11984,10	48,35
	$n = 2$	37,63 (0,000)	1017,59	86,48
	$n = 3$	63,43 (0,000)	1972,90	94,35
	(5)	62,30 (0,000)	–1948,29	82,25
	(6)	64,23 (0,000)	–751,91	88,36

a Zmienna nieistotna statystycznie.

TABL. 4. WSPÓŁCZYNNIKI REGRESJI LINIOWEJ WSKAŹNIKÓW SKOLARYZACJI PODSTAWOWEJ, ŚREDNIEJ I WYŻSZEJ (dok.)

Status krajów	Typ równania	Współczynnik zmiennej niezależnej (t–1), (p-*value*)	Współczynnik zmiennej stałej	R^2 w %
*S*4	(4): n = 1	–663,05[a] (0,253)	85601,61	11,70
	n = 2	293,76 (0,000)	–10272,10	82,79
	n = 3	223,38 (0,000)	9428,14	96,32
	(5)	690,61 (0,000)	–50576,30	85,52
	(6)	420,05 (0,000)	–15165,60	95,05

a Zmienna nieistotna statystycznie.
Ź r ó d ł o: jak przy tabl. 2.

Bez względu na wybór metody obliczania współczynnika skolaryzacji dla wszystkich grup dochodowych uzyskano dodatni współczynnik przy zmiennej niezależnej[6]. Zapewnienie dzieciom i młodzieży dostępu do wszystkich poziomów edukacji wpływa pozytywnie na wielkość PKB *per capita*.

W literaturze przedmiotu panuje pełna zgoda co do pozytywnego wpływu edukacji (lub ogólniej — kapitału ludzkiego) na wzrost gospodarczy. Udowodnili to m.in. Barro (1991), Seebens i Wobst (2003), Mankiw, Rome i Weil (1992).

Bezpośrednie inwestycje zagraniczne

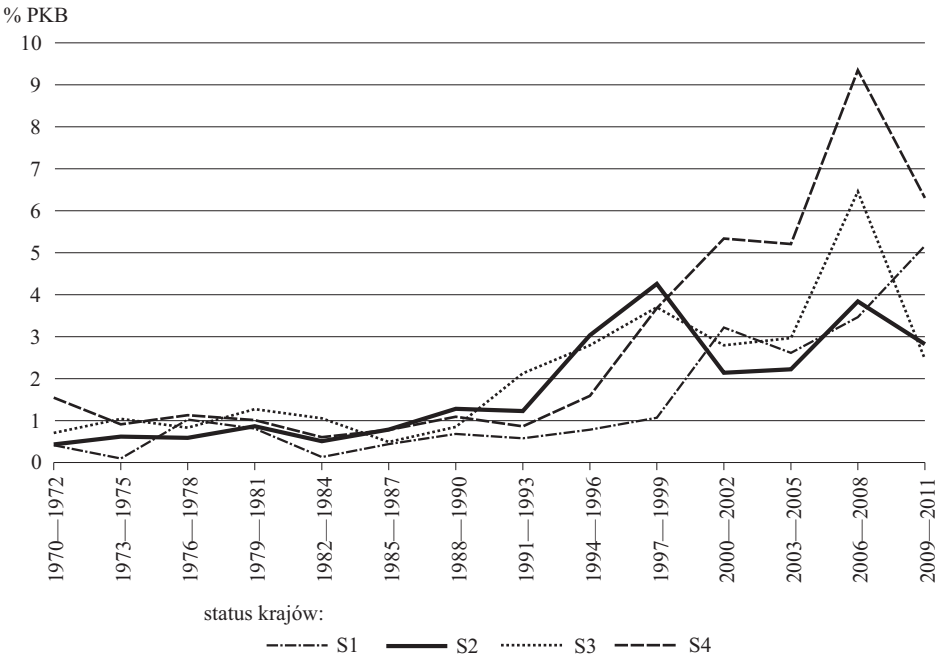
BIZ są to przepływy netto inwestycji przeznaczonych na nabycie trwałego udziału w przedsiębiorstwie[7], które działa w gospodarce innej niż inwestora. BIZ jest sumą kapitału, reinwestycji zysków oraz innych kapitałów krótko- i długoterminowych wykazanych w bilansie płatniczym.

Stosunek BIZ do PKB wzrósł we wszystkich grupach dochodowych w badanym okresie. Obecnie najwyższe wartości notowane są w krajach o najwyższym dochodzie, najniższe zaś w grupie krajów o średniowysokim PKB *per capita*. Z kolei zdecydowany wzrost notowano we wszystkich grupach po 1990 r. W latach kryzysu (2009—2011) można zaobserwować zmniejszenie udziału bezpośrednich inwestycji zagranicznych w PKB.

[6] Wyjątek stanowi współczynnik zmiennej zależnej *EDU*1 w równaniu (3) dla grupy krajów o najwyższym PKB *per capita*, jednak jest on nieistotny statystycznie, co wyklucza wyciągnięcie wniosków.

[7] Trwały udział w przedsiębiorstwie rozumiany jako posiadanie co najmniej 10% akcji z prawem głosu.

Wykr. 5. BEZPOŚREDNIE INWESTYCJE ZAGRANICZNE



Ź r ó d ł o: jak przy wykr. 1.

TABL. 5. WSPÓŁCZYNNIKI REGRESJI LINIOWEJ ZMIENNEJ „BEZPOŚREDNIE INWESTYCJE ZAGRANICZNE"

Status krajów	Współczynnik zmiennej niezależnej (t–1), (p-value)	Współczynnik zmiennej stałej	R^2 w %
*S*1	6,41[a] (0,089)	290,21	24,04
*S*2	105,44 (0,001)	804,58	63,17
*S*3	353,37 (0,000)	2459,36	82,41
*S*4	1224,76 (0,001)	14447,99	64,80

a Zmienna nieistotna statystycznie.
Ź r ó d ł o: jak przy tabl. 2.

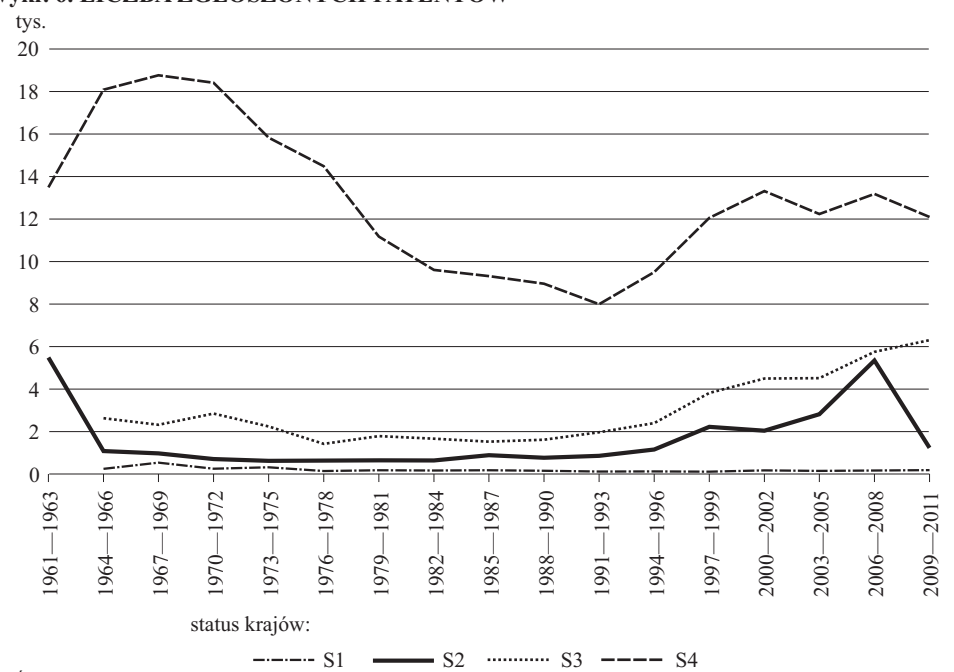
Estymując równania regresji otrzymano dodatnie współczynniki przy zmiennej niezależnej bez względu na grupę dochodową, co oznacza, że BIZ są pozytywną determinantą wzrostu gospodarczego. Należy jednak zaznaczyć, iż równanie dla grupy krajów o najniższym PKB *per capita* jest nieistotne statystycznie.

BIZ wpływają pozytywnie na wzrost gospodarczy pod warunkiem, że w kraju będącym odbiorcą występuje wystarczający poziom kapitału ludzkiego zdolnego do adaptacji nowych technologii. Może występować efekt wypierania (*crowding--out effect*), jeśli inwestycje zagraniczne są komplementarne z krajowymi, do których wpływają (Borensztein, De Gregorio, Lee, 1997). Nair-Reichert i Weinhold (2001) udowodnili pozytywny wpływ BIZ na wzrost gospodarczy oraz zależność polegającą na tym, że im bardziej otwarta jest gospodarka, tym owy wpływ jest większy. W krajach rozwijających się BIZ mają większy wpływ na wzrost gospodarczy niż inwestycje krajowe, również w długim okresie (Hansen, Rand, 2006).

Innowacyjność

Kolejną badaną determinantą wzrostu gospodarczego jest innowacyjność gospodarki. Definiowana jest ona w zaprezentowanej analizie jako liczba złożonych wniosków patentowych w krajowych urzędach patentowych oraz w innych miejscach na świecie za pomocą międzynarodowego zgłoszenia PTC (*Patent Cooperation Treaty*) przez rezydentów danego kraju.

Wykr. 6. LICZBA ZGŁOSZONYCH PATENTÓW



Ź r ó d ł o: jak przy wykr. 1.

Kraje o najwyższym dochodzie charakteryzują się największą liczbą zgłoszonych wniosków patentowych. Wraz z przechodzeniem do grup o coraz niższym PKB *per capita* następuje spadek badanej zmiennej.

TABL. 6. WSPÓŁCZYNNIKI REGRESJI LINIOWEJ ZMIENNEJ „PATENTY"

Status krajów	Współczynnik zmiennej niezależnej (t–1), (p-*value*)	Współczynnik zmiennej stałej	R^2 w %
*S*1	0,05[a] (0,178)	288,60	13,51
*S*2	0,12 (0,000)	765,05	68,03
*S*3	0,42 (0,001)	1884,91	59,43
*S*4	–0,67[a] (0,036)	25271,84	29,52

a Zmienna nieistotna statystycznie.
Ź r ó d ł o: jak przy tabl. 2.

Otrzymano dwa równania regresji istotne statystycznie dla krajów o średnioniskim i średniowysokim dochodzie. W obu przypadkach uzyskano potwierdzenie, że innowacyjność gospodarki (mierzona liczbą złożonych wniosków patentowych) powiększa PKB *per capita*.

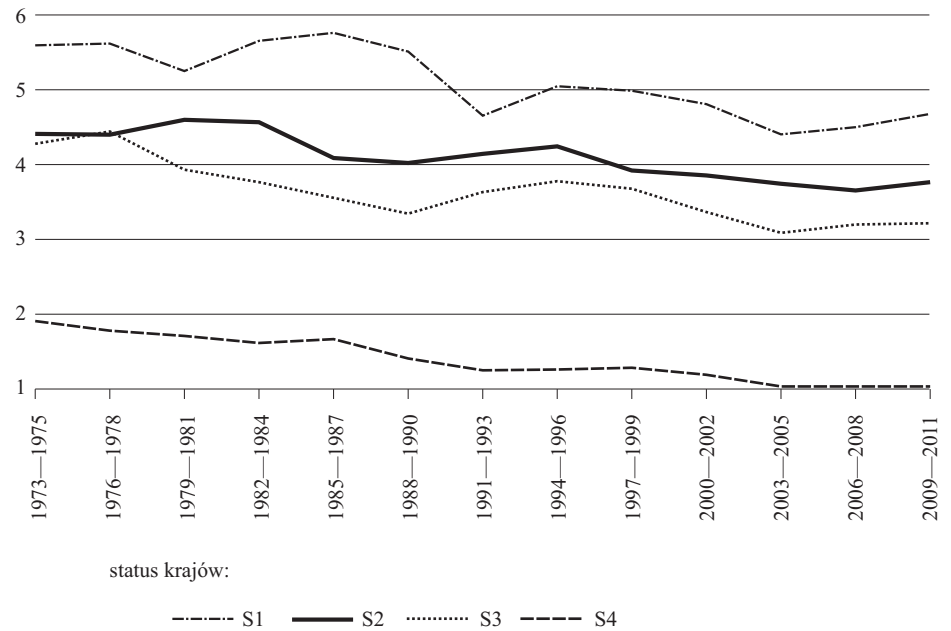
W literaturze za „innowacyjność", obok liczby złożonych patentów, przyjmuje się także wydatki w sektorze badań i rozwoju (*R&D — Research and Development*). Panuje niepodzielna zgoda, że innowacyjność gospodarki sprzyja uzyskiwaniu dodatniego tempa wzrostu PKB *per capita*. Zależność tę udowodnił Romer (1986) w endogenicznym modelu *learning-by-doing* (uczenie się przez pracę). Identyczne relacje innowacji i wzrostu gospodarczego przedstawili Grossman i Helpman (1990) oraz Durlauf, Johnson i Temple (2005).

Wolność osobista i polityczna

Wskaźnik wolności FreedomHouse jest średnią arytmetyczną wskaźników wolności osobistej oraz praw politycznych. Wskaźnik wolności osobistej określa stopień wolności słowa i przekonań, prawa do stowarzyszania się, praworządności, niezależności oraz braku ingerencji ze strony państwa we wskazanych dziedzinach. Wskaźnik praw politycznych dotyczy nieskrępowanego uczestnictwa w procesach politycznych, w tym prawa do swobodnego głosowania w wolnych i legalnych wyborach, rywalizacji o urzędy publiczne, przynależności do partii politycznych i organizacji oraz wybierania przedstawicieli mających faktyczny i decydujący wpływ na prowadzoną politykę.

Organizacja FreedomHouse nadaje każdemu z tych dwóch wskaźników wartość z przedziału od 1 do 7, przy czym 7 oznacza całkowity brak wolności (lub praw politycznych), zaś 1 — całkowitą wolność.

Wykr. 7. WSKAŹNIK WOLNOŚCI FreedomHouse



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

We wszystkich badanych latach najwyższym poziomem wolności charakteryzowały się kraje o największym dochodzie, najniższym — kraje najbiedniejsze. Bez względu na przynależność do grupy dochodowej w ostatnich latach obserwowana jest stopniowa poprawa wolności i praw politycznych.

TABL. 7. WSPÓŁCZYNNIKI REGRESJI LINIOWEJ ZMIENNEJ „WSKAŹNIK WOLNOŚCI FreedomHouse"

Status krajów	Współczynnik zmiennej niezależnej (t–1), (p-*value*)	Współczynnik zmiennej stałej	R^2 w %
*S*1	–5,81[a] (0,568)	326,75	3,37
*S*2	–468,90 (0,000)	2939,95	76,01
*S*3	–1142,59 (0,008)	7452,52	52,54
*S*4	–12218,6 (0,000)	35498,75	91,45

a Zmienna nieistotna statystycznie.
Ź r ó d ł o: opracowanie własne na podstawie danych bazy FreedomHouse.

Na podstawie estymacji KMNK otrzymano trzy istotne statystycznie równania. Jedynie w przypadku grupy krajów o najniższym dochodzie współczynnik zmiennej niezależnej charakteryzował się nieistotnością. Ujemne wartości

współczynników oznaczają, że im więcej wolności i praw politycznych w danym kraju, tym PKB *per capita* wyższy.

W literaturze istnieje wiele terminów opisujących instytucje państwa, jednak bez względu na przyjęte definicje wskazuje się jednoznacznie na ich pozytywny wpływ na gospodarkę. Znalazło to odzwierciedlenie w badaniach dotyczących oddziaływania rządów prawa na wzrost gospodarczy (Knack, Keefer, 1995), a także w analizach wpływu sprawności instytucji (jakości rządzenia) na wzrost i poziom produkcji (Rivera-Batiz, 2002; Tridico, 2006).

TEST PRZYCZYNOWOŚCI

Wnioski z przeprowadzonych analiz obrazuje tabl. 8.

TABL. 8. WYNIKI TESTU PRZYCZYNOWOŚCI W SENSIE GRANGERA WYBRANYCH ZMIENNYCH NA PKB *PER CAPITA*

Zmienne niezależne	Liczba opóźnień		
	1	2	3
Wydatki rządowe	x	►	►
Inflacja	►	►	x
Współczynnik solaryzacji	x	x	x
BIZ	x	►	►
Patenty złożone	x	x	►
Wskaźnik wolności FreedomHouse	►	x	►

U w ag a. „►" oznacza, że analizowana zmienna niezależna jest przyczyną w sensie Grangera poziomu PKB *per capita*.
Ź r ó d ł o: opracowanie własne.

W tabl. 9 zaprezentowano wyniki testu Walda użytego do określenia przyczynowości.

Na podstawie przeprowadzonego testu przyczynowości można stwierdzić, że wszystkie badane zmienne (poza współczynnikiem skolaryzacji) są przyczyną w sensie Grangera poziomu PKB *per capita*. Wydatki rządowe oraz BIZ wpływają na zmienną zależną po dwóch okresach (6 lat), inflacja i wskaźnik wolności FreedomHouse już po jednym okresie (3 lata), zaś liczba złożonych patentów po trzech okresach (9 lat).

TABL. 9. WYNIKI TESTU WALDA DLA HIPOTEZY ZEROWEJ[a]

Zmienne	Opóźnienia								
	1			2			3		
	liczba obserwacji	wartość		liczba obserwacji	wartość		liczba obserwacji	wartość	
		empiryczna	krytyczna		empiryczna	krytyczna		empiryczna	krytyczna
Wydatki rządowe	898	–201,57	3,84	836	84,48	5,99	774	26,16	5,99
Inflacja	803	177,83	3,84	738	9,08	5,99	676	–65,83	5,99
Współczynnik skolaryzacji	672	–179,65	3,84	580	1,39	5,99	494	–24,16	5,99

a Badana zmienna nie jest przyczyną w sensie Grangera PKB *per capita*.

TABL. 9. WYNIKI TESTU WALDA DLA HIPOTEZY ZEROWEJ[a] (dok.)

Zmienne	Opóźnienia								
	1			2			3		
	liczba obserwacji	wartość		liczba obserwacji	wartość		liczba obserwacji	wartość	
		empiryczna	krytyczna		empiryczna	krytyczna		empiryczna	krytyczna
BIZ	646	–129,75	3,84	564	96,35	5,99	488	82,11	5,99
Patenty złożone	559	–52,66	3,84	485	–150,46	5,99	420	42,90	5,99
Wskaźnik wolności FreedomHouse	684	12,66	3,84	626	–26,45	5,99	568	23,73	5,99

a Badana zmienna nie jest przyczyną w sensie Grangera PKB *per capita*.

Ź r ó d ł o: opracowanie własne.

Zakończenie

W artykule przedstawiono wybrane determinanty wzrostu gospodarczego wraz z kształtowaniem się ich wartości w latach 1961—2011. Utworzone modele regresji liniowej pozwoliły na określenie pozytywnego wpływu wydatków rządowych, skolaryzacji, BIZ oraz innowacyjności danej gospodarki na PKB *per capita*. Zbadano, że negatywnie na dochód liczony na osobę wpływała inflacja, choć wnioskowanie w tym przypadku jest utrudnione. Potwierdziło się również i to, że wolność osobista i prawa polityczne sprzyjają wyższemu poziomowi PKB *per capita*.

Test przyczynowości w sensie Grangera wykazał, że za poziom PKB *per capita* odpowiadają: wydatki rządowe, inflacja, BIZ, innowacyjność, a także wolność osobista i polityczna. Potwierdzono zatem hipotezę postawioną we wstępie opracowania.

W wyniku przeprowadzonej analizy stwierdzono, że w dalszych badaniach tego zagadnienia należałoby wykorzystać inne metody określenia wpływu zmiennych niezależnych na PKB *per capita*, np. poprzez badania panelowe. Uzasadnione byłoby też estymowanie równań regresji liniowej badających wpływ zmiennej niezależnej o większej liczbie opóźnień (w tym opracowaniu przyjęto wyłącznie opóźnienia o 1 okres). Zasadne byłoby uzupełnienie braku danych lub uzyskanie ich z innych źródeł, aby uniknąć nieścisłości i przekłamań na etapie gromadzenia danych. Zwiększenie liczby badanych krajów oraz liczby zmiennych mogłoby pozwolić na uzyskanie bardziej wyczerpujących wniosków.

mgr Bartosz Totleben — *Uniwersytet Ekonomiczny w Poznaniu*

LITERATURA

Al. Bataineh I. M. (2012), *The impact of government expenditures on economic growth in Jordan*, Interdisciplinary Journal of Contemporary Research in Business

Barro R. (1990), *Government Spending in a Simple Model of Endogenous Growth*, „The Journal of Political Economy", Vol. 98, No. 5, Part 2

Barro R. (1991), *Economic growth in a cross-section of countries*, „The Quarterly Journal of Economics"

Blundell R., Bond S. (1998), *Initial conditions and moment restrictions in dynamic panel data models*, „Journal of Econometrics", Vol. 87

Borensztein E., De Gregorio J., Lee J-W. (1997), *How does foreign direct investment affect economic growth?*, „Journal of International Economics", No. 45

Bruno M., Easterly W. (1995), *Inflation crises and long-run growth*, National Bureau of Economic Research, Cambridge

Chowdhury A., Mallik G. (2001), *Inflation and Economic Growth: Evidence From Four South Asian Countries*, Asia-Pacific Development Journal

Durlauf S., Johnson P., Temple J. (2005), *Growth econometrics* [w: *Handbook of Economic Growth,* red. Aghion P., Durlauf S., Amsterdam

Granger C. (1969), *Investigating causal relations by econometric models and cross-spectral methods*, „Econometrica", Vol. 37(3)

Grossman G., Helpman E. (1990), *Trade, Innovation, and Growth*, „The American Economic Review", Vol. 80, No. 2

Hansen H., Rand J. (2006), *On the Causal Links Between FDI and Growth in Developing Countries*, „The World Economy", Vol. 29(1)

Jardat M. A. (2013), *Impact of inflation and unemployment on Jordanian GDP*, Interdisciplinary Journal of Contemporary Research in Business

Knack S., Keefer P. (1995), *Institutions and Economic Performance: Cross Country Tests Using Alternative Institutional Measures*, University of Maryland, Center for Institutional Reform and the Informal Sector

Kormendi R., Meguire P. (1985), *Macroeconomic determinants of growth: Cross-country evidence*, „Journal of Monetary Economics", Vol. 16

Landau D. (1983), *Government Expenditure and Economic Growth: A Cross-Country Study*, „Southern Economic Journal", Vol. 49, No. 3

Mankiw N., Romer D., Weil D. (1992), *A Contribution to the Empirics of Economic Growth*, The Quarterly Journal of Economics

Nair-Reichert U., Weinhold D. (2001), *Causality tests for cross-country panels: a new look at FDI and economic growth in developing countries*, „Oxford Bulletin of Economics and Statistic", No. 63

Pan Y. (2013), *Study on the Effect of Goverment Spending on GDP Growth*, [w:] *Informatics and Management Science VI*, Springer, London

Piątek D., Pilc M., Szarzec K. (2013), *Economic freedom, democracy and economic growth: a causal investigation in transition countries*, „Post-Communist Economies", Vol. 25, No. 3

Rivera-Batiz F. (2002), *Democracy, Governance and Economic Growth: Theory and Evidence*, „Review of Development Economics", Vol. 6, No. 2

Romer P. (1986), *Increasing Returns and Long-Run Growth*, „The Journal of Political Economy", Vol. 93, No. 5

Sala-i-Martin X. (2013), *Some Lessons From 10 years of Empirical Growth Literature*, http://www.clmeconomia.jccm.es/pdfclm/xavier_i.pdf (15.12.2013)

Seebens H., Wobst P. (2003), *The impact of Increased School Enrollment on Economic Growth in Tanzania*, Disscussion Papers on Development Policy, Bonn

Tridico P. (2006), *Institutional Change and Governance Indexes in Transition Economies: The Case of Poland*, „The European Journal of Comparative Economics", Vol. 3, No. 2

Windmeijer F. (2005), *A finite sample correction for the variance of linear efficient two-step GMM estimators*, „Journal of Econometrics", Vol. 126(1)

SUMMARY

The article examines the impact of selected variables such as the level of government spending, inflation, school enrollment, FDI, innovation and economic freedom on GDP per capita level. To do so, linear regression, Granger causality test and general method of moments are employed. The research takes into consideration 59 countries, divided into 4 groups accordingly to GNP. The results confirm stated hypothesis, i.e. selected variables determine economic growth, what coincides with the analysis of literature.

РЕЗЮМЕ

В статье было рассмотрено влияние избранных переменных: уровня правительственных издержек, инфляции, сколяризации, прямых внешних капиталовложений, инноваций и экономической свободы на уровень ВВП per capita. Обследованием были охвачены 59 стран, которые согласно классификации Всемирного банка были разделены на 4 доходные группы. Для определения зависимости и влияния представленных переменных на ВВП per capita были использованы: линейная регрессия, причинный критерий Грангера и обобщенный метод моментов. Результат обследований это подтверждение поствленного гипотеза, что избранные переменные являются детерминантами экономического роста, что согласуется с представленным анализом литературы.

Anna TURCZAK, Patrycja ZWIECH

Zależność między miejscem zamieszkania na wsi i w mieście a głównym źródłem utrzymania ludności

Praca najemna jest wciąż najczęstszym źródłem utrzymania[1] ludności w Polsce. Jednak liczba ludności, dla której dochody z pracy najemnej są głównym źródłem utrzymania nie stanowi nawet połowy liczebności mieszkańców kraju mających własne źródło utrzymania, zatem nie jest jedynym ważnym źródłem utrzymania Polaków. W 2011 r. głównie z pracy najemnej utrzymywało się 48,9% osób mających własne źródło dochodów, z pracy na własny rachunek poza rolnictwem — 6,1%, z pracy na własny rachunek w rolnictwie — 4,8%, ze źródeł niezarobkowych — 40,0%, a z własności — niecałe 0,2%. Z kolei

[1] Przy przeprowadzaniu badań Autorki posłużyły się klasyfikacją źródeł utrzymania podaną przez GUS. Zgodnie z tą klasyfikacją:

- **praca najemna** obejmuje zarówno pracę świadczoną w sektorze publicznym, jak i pracę świadczoną w sektorze prywatnym;
- **praca na rachunek własny poza rolnictwem** to prowadzenie własnej działalności gospodarczej albo wykonywanie wolnego zawodu. W opisywanej grupie znajdują się również osoby prowadzące jednoosobowy podmiot gospodarczy, którym to osobom narzucono zarejestrowanie własnej działalności gospodarczej. Co istotne, także dochody z wynajmu są traktowane w wykorzystywanej tutaj klasyfikacji jako pochodzące z prowadzonej działalności gospodarczej, stąd osoby, dla których głównym źródłem dochodu jest wynajem, w prezentowanych zestawieniach nie zostały ujęte oddzielnie, lecz razem z pracującymi na własny rachunek poza rolnictwem;
- **praca na rachunek własny w rolnictwie** to praca w swoim gospodarstwie rolnym związana z produkcją roślinną i/lub zwierzęcą. Do grupy tej zaliczono również członków rolniczych spółdzielni produkcyjnych, kółek rolniczych oraz członków spółek cywilnych w rolnictwie. Jako praca na własny rachunek w rolnictwie traktowana jest również praca wykonywana przez osoby utrzymujące się z prowadzenia działalności usługowej związanej z rolnictwem (czyli taka działalność, jak: nawożenie pól, opryski upraw, obsługa systemów irygacyjnych, sztuczne unasiennianie zwierząt, usługi pasterskie, usługi w zakresie leśnictwa lub łowiectwa);
- **niezarobkowe źródło dochodu** to emerytura (pracownicza, kombatancka, rolna), renta: strukturalna, z tytułu niezdolności do pracy, inwalidzka, rodzinna (wdowia, sieroca) i socjalna, zasiłek dla bezrobotnych, świadczenia i zasiłki przedemerytalne oraz świadczenia pomocy społecznej;
- **dochody z własności** to dochody z tytułu oddanych w dzierżawę gruntów rolnych, dochody z lokat kapitałowych (z obrotu akcjami, obligacjami, z zysków kapitałowych itp.), z odsetek od oszczędności i z udziału w zyskach przedsiębiorstw (dywidend) — *Ludność…* (2013), s. 22—24.

Dodatkowo GUS w klasyfikacji podaje też **inne źródła niż wymienione** (do kategorii tej należą m.in. alimenty od osób prywatnych, stypendia naukowe i sportowe), jednak znaczenie tej pozycji — w przypadku rozpatrywania w niniejszym artykule jedynie osób posiadających własne źródło utrzymania — jest znikome i dlatego też we wszystkich przeprowadzonych analizach zostało pominięte.

w 2002 r. głównie z pracy najemnej utrzymywało się w Polsce 41,7% osób mających własne źródło dochodów, z pracy na własny rachunek poza rolnictwem — 5,5%, z pracy na własny rachunek w rolnictwie — 6,4%, ze źródeł niezarobkowych — 46,3%, a z własności — jedynie nieco ponad 0,1%. Z porównania przywołanych danych za 2011 r. z odpowiednimi danymi dotyczącymi 2002 r. można wyciągnąć następujące wnioski[2]:

- znacznie wzrósł udział osób, dla których dochody z pracy najemnej są głównym źródłem utrzymania (wzrost o 7,2 p.proc.);
- wzrósł udział osób, w przypadku których dochody z pracy na własny rachunek poza rolnictwem są głównym źródłem utrzymania (wzrost o 0,6 p.proc.), a spadł udział osób, dla których dochody z pracy na własny rachunek w rolnictwie są głównym źródłem utrzymania (spadek o 1,6 p.proc.). Dzięki tym zmianom w 2011 r. więcej mieszkańców Polski czerpało główne dochody z pracy na własny rachunek poza rolnictwem niż z pracy na własny rachunek w rolnictwie, podczas gdy w 2002 r. było na odwrót;
- udział osób, w przypadku których dochody ze źródeł niezarobkowych są głównym źródłem utrzymania spadł o 6,3 p.proc.;
- z własności swoje główne dochody pobierało w 2011 r. więcej osób niż w 2002 r. i liczebność tej grupy rosła najbardziej dynamicznie (liczba takich osób zwiększyła się na przestrzeni dziewięciu lat o 32%). Jednak udział osób, dla których dochody z własności były głównym źródłem utrzymania w 2002 r. był bardzo niski, z tego względu — pomimo dynamicznego wzrostu na przestrzeni badanego okresu — udział ten w 2011 r. nadal pozostał relatywnie niski i zwiększył się jedynie o część p.proc.

Na to, jak kształtuje się struktura ludności pod względem głównego źródła utrzymania wpływ ma wiele różnorodnych czynników[3]. Czynnikiem, który kształtuje strukturę ludności Polski — w przypadku rozpatrywania głównego źródła utrzymania Polaków jako cechy będącej przedmiotem badania — jest także miejsce zamieszkania. I właśnie odpowiedź na pytanie, w jakim stopniu fakt, że osoba mieszka na wsi lub w mieście wpływa na rodzaj jej głównego źró-

[2] Wszystkie wnioski oparto na danych pochodzących z Narodowego Spisu Powszechnego Ludności i Mieszkań 2002 oraz Narodowego Spisu Powszechnego Ludności i Mieszkań 2011, których wyniki Główny Urząd Statystyczny udostępnił w publikacjach *Ludność*... (2003), s. 62—79 oraz *Ludność*... (2013), s. 92—97.

[3] Takimi czynnikami są przykładowo wiek i wykształcenie. Zdecydowanie większy jest udział osób utrzymujących się z pracy najemnej w przypadku trzydziesto- czy czterdziestolatków niż wśród siedemdziesięcio- i osiemdziesięciolatków. Z kolei udział osób utrzymujących się ze źródeł niezarobkowych jest większy wśród siedemdziesięcio- i osiemdziesięciolatków niż w grupie trzydziesto- czy czterdziestolatków — *Ludność*... (2013), s. 32—34. Czynnikiem, który ma wpływ na strukturę społeczeństwa pod względem głównego źródła utrzymania jest również wykształcenie. Okazuje się, że osoby ze średnim i wyższym wykształceniem relatywnie częściej czerpią swoje główne dochody z pracy najemnej bądź z pracy na własny rachunek poza rolnictwem niż osoby z wykształceniem zasadniczym zawodowym, gimnazjalnym czy podstawowym. Z kolei dla osób z wykształceniem podstawowym, gimnazjalnym i zasadniczym zawodowym stosunkowo częściej źródła niezarobkowe są głównym źródłem utrzymania — *Ludność*... (2013), s. 36—39.

dła dochodu, stała się celem obecnego opracowania. W związku z celem pracy postawiono hipotezę, że fakt mieszkania w mieście albo na wsi w coraz mniejszym stopniu determinuje rozkład mieszkańców Polski pod względem rodzajów posiadanych źródeł utrzymania. Aby udzielić pełnych odpowiedzi dotyczących tych kwestii, przeprowadzono trzyetapowe badanie, które stanowi kolejne podrozdziały tego opracowania. Badaniem objęto następujące zagadnienia:

1) określenie związku między miejscem zamieszkania a głównym źródłem utrzymania mieszkańców Polski w 2002 r.;
2) określenie związku między miejscem zamieszkania a głównym źródłem utrzymania mieszkańców Polski w 2011 r.;
3) porównanie siły zależności między miejscem zamieszkania a głównym źródłem utrzymania mieszkańców Polski w latach 2002 i 2011.

W celu sprawdzenia, czy występuje statystycznie istotny związek między rozpatrywanymi cechami wykorzystano test niezależności *chi*-kwadrat przeprowadzony dla lat 2002 i 2011. Z kolei do określenia siły zależności występujących w poszczególnych latach użyto współczynnika *V* Cramera. W ostatnim etapie — na podstawie porównania wartości współczynnika *V* Cramera obliczonego dla 2002 r. z jego wartością uzyskaną dla 2011 r. — wyciągnięto wnioski pozwalające na całkowitą realizację głównego celu badania.

ISTOTA ZASTOSOWANEGO NARZĘDZIA BADAWCZEGO

Test niezależności *chi*-kwadrat jest nieparametrycznym testem istotności, który może być wykorzystywany do oceny zależności stochastycznej dwóch cech jakościowych, dwóch cech ilościowych, a także dowolnej cechy ilościowej i jakościowej. Punktem wyjścia do przeprowadzenia testu jest sporządzenie tablicy, której wnętrze stanowi liczebność empiryczna, czyli zaobserwowana. Tablica ta jest macierzą o r wierszach i s kolumnach, przy czym r oznacza liczbę wariantów pierwszej cechy (X), a s — liczbę wariantów drugiej cechy (Y). Z kolei n_{ij} (i=1, 2, …, r, j=1, 2, …, s) to liczba tych obserwacji, dla których cecha X przyjmuje wariant x_i, a cecha Y — wariant y_j.

Sumując osobno każdy wiersz i każdą kolumnę macierzy liczebności empirycznej otrzymuje się tzw. liczebność brzegową, którą oznaczono jako $n_{i\cdot}$ i $n_{\cdot j}$. Zachodzą więc równości[4]:

$$n_{i\cdot} = \sum_{j=1}^{s} n_{ij} \qquad n_{\cdot j} = \sum_{i=1}^{r} n_{ij} \tag{1}$$

Czyli $n_{i\cdot}$ to liczebność brzegowa w wierszu o numerze i, którą uzyskano dodając każdą liczebność znajdującą się w tym wierszu, natomiast $n_{\cdot j}$ to liczebność brzegowa w kolumnie o numerze j, którą obliczono dodając każdą liczeb-

[4] Kot i in. (2007), s. 294 i 295.

ność leżącą w tej kolumnie. Poprawność wykonanych obliczeń można skontrolować sprawdzając, czy suma liczebności brzegowej dotycząca wierszy jest taka sama, jak suma liczebności brzegowej z kolumn i równa się liczebności całkowitej n. Tak więc:

$$n = \sum_{i=1}^{r}\sum_{j=1}^{s} n_{ij} = \sum_{i=1}^{r} n_{i\cdot} = \sum_{j=1}^{s} n_{\cdot j} \tag{2}$$

Tabl. 1 przedstawia schemat macierzy liczebności empirycznej.

TABL. 1. SCHEMAT MACIERZY LICZEBNOŚCI EMPIRYCZNEJ

Wyszczególnienie		Warianty drugiej cechy (Y)						$n_{i\cdot}$
		y_1	y_2	...	y_j	...	y_s	
Warianty pierwszej cechy (X)	x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1s}	$n_{1\cdot}$
	x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2s}	$n_{2\cdot}$
	...	...	...	...	...	...	...	...
	x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{is}	$n_{i\cdot}$
	...	...	...	...	...	...	...	...
	x_r	n_{r1}	n_{r2}	...	n_{rj}	...	n_{rs}	$n_{r\cdot}$
$n_{\cdot j}$		$n_{\cdot 1}$	$n_{\cdot 2}$	...	$n_{\cdot j}$	...	$n_{\cdot s}$	n

Ź r ó d ł o: opracowanie własne, gdzie: r — liczba wariantów cechy X, s — liczba wariantów cechy Y, n_{ij} — liczba obserwacji posiadających i-ty wariant cechy X oraz j-ty wariant cechy Y.

Mając tak przygotowaną macierz liczebności empirycznej można sformułować hipotezę zerową H_0, iż badane dwie cechy są stochastycznie niezależne wobec hipotezy alternatywnej H_1, ponieważ występuje stochastyczna zależność między tymi cechami[5]. Aby sprawdzić prawdziwość hipotezy zerowej, na podstawie obliczonej liczebności brzegowej i liczebności całkowitej należy wyznaczyć prawdopodobieństwa brzegowe zgodnie z formułami[6]:

$$p_{i\cdot} = \frac{n_{i\cdot}}{n} \qquad p_{\cdot j} = \frac{n_{\cdot j}}{n} \tag{3}$$

gdzie:

$p_{i\cdot}$ — prawdopodobieństwo, że obserwacja przyjmuje i-ty wariant cechy X,

$p_{\cdot j}$ — prawdopodobieństwo, że obserwacja przyjmuje j-ty wariant cechy Y.

[5] Aczel (2000), s. 758.

[6] Kukuła (2003), s. 196.

Z kolei prawdopodobieństwa empiryczne wewnątrz tablicy można obliczyć według wzoru:

$$p_{ij} = \frac{n_{ij}}{n} \tag{4}$$

gdzie p_{ij} — prawdopodobieństwo, że obserwacja przyjmuje *i*-ty wariant cechy *X* i *j*-ty wariant cechy *Y*.

Następnie, mnożąc kolejne prawdopodobieństwa brzegowe dotyczące wierszy przez prawdopodobieństwa brzegowe z poszczególnych kolumn otrzymuje się macierz prawdopodobieństw teoretycznych p_{ij}^*. Obliczone wielkości są hipotetycznymi prawdopodobieństwami, które wystąpiłyby, gdyby hipoteza zerowa była prawdziwa, czyli gdyby rozpatrywane cechy *X* i *Y* były niezależne. Zatem prawdopodobieństwa teoretyczne wyznacza się zgodnie ze wzorem[7]:

$$p_{ij}^* = p_{i\cdot} p_{\cdot j} \text{ przy czym } \sum_{i=1}^{r}\sum_{j=1}^{s} p_{ij}^* = 1, \text{ czyli } 100\% \tag{5}$$

Treść hipotezy zerowej i hipotezy alternatywnej można formalnie przedstawić następująco[8]:

$$H_0\colon\ E\left(p_{ij} - p_{ij}^*\right)^2 = 0 \text{ przeciwko } H_1\colon\ E\left(p_{ij} - p_{ij}^*\right)^2 > 0 \tag{6}$$

gdzie symbol *E* oznacza wartość oczekiwaną.

W następnym etapie należy obliczyć liczebność hipotetyczną n_{ij}^*, która wystąpiłaby, gdyby spełniony był warunek o niezależności cech. Poszczególne wartości n_{ij}^* otrzymano mnożąc odpowiednie prawdopodobieństwa teoretyczne p_{ij}^* przez liczebność całkowitą n, a więc postępując według wzoru[9]:

$$n_{ij}^* = np_{ij}^* \tag{7}$$

[7] Ostasiewicz i in. (1995), s. 264.
[8] Podgórski (2005), s. 274.
[9] Zeliaś (2000), s. 286.

Ostatecznie na podstawie porównania elementów macierzy liczebności rzeczywistej n_{ij} z elementami macierzy liczebności teoretycznej n_{ij}^* należy zdecydować, czy odrzucić hipotezę H_0 na skutek wystąpienia dużych różnic między tymi dwoma rodzajami liczebności czy jednak nie ma podstaw do odrzucenia hipotezy zerowej. Podjęcie decyzji ułatwia zastosowanie statystyki χ^2, którą przyjmuje się jako syntetyczną miarę odchyleń liczebności rzeczywistej od liczebności teoretycznej. Wartość statystyki testowej χ^2 obliczono zgodnie z formułą[10]:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\left(n_{ij} - n_{ij}^*\right)^2}{n_{ij}} \tag{8}$$

Statystyka χ^2, przy założeniu prawdziwości hipotezy H_0 o niezależności cech, ma asymptotyczny rozkład *chi*-kwadrat o $(r-1)(s-1)$ stopniach swobody. Przyjmuje ona wartości z przedziału $\langle 0, n\cdot\min(r-1)(s-1)\rangle$. Z jej budowy wynika, że im większe są rozbieżności między liczebnością empiryczną n_{ij} i oczekiwaną n_{ij}^*, tym wyższa jest wartość obliczonej statystyki χ^2. Statystyka χ^2 jest równa zeru, gdy poszczególna liczebność empiryczna i teoretyczna są takie same, a zatem rozpatrywane cechy są stochastycznie niezależne. Z kolei maksymalną wartość, tj. $n\cdot\min(r-1)(s-1)$, przyjmuje ona w przypadku zależności funkcyjnej między cechami X i Y[11].

Przy podejmowaniu decyzji w teście niezależności *chi*-kwadrat należy brać pod uwagę jednostronny (a konkretnie — prawostronny) obszar krytyczny, który określa nierówność $\chi^2 \geq \chi_\alpha^2$, gdzie χ_α^2 jest wartością krytyczną odczytaną z tablic rozkładu *chi*-kwadrat dla przyjętego z góry poziomu istotności α i dla $(r-1)(s-1)$ stopni swobody w taki sposób, aby zachodziła relacja $P\{\chi^2 \geq \chi_\alpha^2\} = \alpha$. Obliczoną wartość statystyki testowej χ^2 porównuje się z wartością krytyczną χ_α^2 i jeśli spełniona jest nierówność $\chi^2 \geq \chi_\alpha^2$, to przy przyjętym poziomie istotności hipotezę H_0 należy odrzucić na korzyść hipotezy H_1. Oznacza to, że rozpatrywane cechy są zależne. Gdy natomiast zachodzi nierówność $\chi^2 < \chi_\alpha^2$, nie ma podstaw do odrzucenia hipotezy zerowej o niezależności badanych cech.

Do określenia siły zależności między cechami wykorzystano współczynnik V Cramera. Współczynnik ten jest wielkością niemianowaną i unormowaną

[10] Krysicki i in. (2003), s. 104.
[11] Jóźwiak, Podgórski (1995), s. 233.

— przyjmuje wartości wyłącznie z przedziału $\langle 0,1\rangle$ [12]. Jeżeli wynosi 0, to między cechami nie występuje zależność. Z kolei im jego wartość jest bliższa 1, tym zależność jest silniejsza. Współczynnik V Cramera obliczono według wzoru[13]:

$$V = \sqrt{\frac{\chi^2}{n \cdot \min(r-1)(s-1)}} \tag{9}$$

ANALIZA WPŁYWU MIEJSCA ZAMIESZKANIA NA ROZKŁAD GŁÓWNYCH ŹRÓDEŁ UTRZYMANIA LUDNOŚCI W POLSCE W 2002 R.

Pierwszym postawionym przed Autorkami zadaniem jest odpowiedź na pytanie, czy w 2002 r. istniała zależność między tym, że osoba mieszka w mieście czy na wsi a rodzajem jej głównego źródła utrzymania. A zatem weryfikacji podlega hipoteza H_0 stanowiąca, że badane dwie cechy są stochastycznie niezależne wobec hipotezy alternatywnej H_1 orzekającej, że występuje stochastyczna zależność między tymi cechami. Procedurę weryfikacyjną zrealizowano za pomocą testu niezależności *chi*-kwadrat. Dane stanowiące punkt wyjścia do przeprowadzenia testu zaprezentowano w tabl. 2[14].

TABL. 2. LICZEBNOŚĆ EMPIRYCZNA (rzeczywista) W 2002 R.

Wyszczególnienie	Główne dochody					Suma
	z pracy			ze źródeł niezarobkowych	z własności	
	najemnej	na własny rachunek poza rolnictwem	na własny rachunek w rolnictwie			
Miasta	6778680	952481	82260	6596832	19841	14430094
Wieś	2836644	311780	1392704	4094853	7229	8643210
Suma	9615324	1264261	1474964	10691685	27070	23073304

Ź r ó d ł o: opracowanie własne na podstawie publikacji *Ludność*... (2003), s. 68 i 74.

Na podstawie informacji prezentowanych w tabl. 2 oraz wykorzystując wzór (3) i wzór (4) wyznaczono prawdopodobieństwa empiryczne (tabl. 3).

[12] Pułaska-Turyna (2005), s. 254.

[13] Buga, Kassyk-Rokicka (2008), s. 121.

[14] Jak już wspomniano, przeprowadzone badanie obejmuje jedynie tych mieszkańców Polski, którzy mają własne źródło utrzymania. Do badanej zbiorowości nie zakwalifikowały się zatem te osoby, które są na utrzymaniu innych osób. Zgodnie z danymi GUS w 2002 r. główne źródło utrzymania w postaci pracy najemnej, pracy na własny rachunek w rolnictwie bądź poza rolnictwem, własności albo źródła niezarobkowego miały 23073304 osoby.

TABL. 3. PRAWDOPODOBIEŃSTWA EMPIRYCZNE (rzeczywiste) W 2002 R. W %

Wyszczególnienie	Główne dochody					Suma
	z pracy			ze źródeł niezarobkowych	z własności	
	najemnej	na własny rachunek poza rolnictwem	na własny rachunek w rolnictwie			
Miasta	29,379	4,128	0,357	28,591	0,086	62,540
Wieś	12,294	1,351	6,036	17,747	0,031	37,460
Suma	41,673	5,479	6,393	46,338	0,117	100,000

Ź r ó d ł o: opracowanie własne na podstawie tabl. 2.

Następnie obliczono prawdopodobieństwa teoretyczne, które występowałyby przy stochastycznej niezależności badanych cech. Do wyznaczenia tych prawdopodobieństw posłużył wzór (5). Otrzymane wyniki prezentuje tabl. 4.

TABL. 4. PRAWDOPODOBIEŃSTWA HIPOTETYCZNE (teoretyczne) W 2002 R. W %

Wyszczególnienie	Główne dochody					Suma
	z pracy			ze źródeł niezarobkowych	z własności	
	najemnej	na własny rachunek poza rolnictwem	na własny rachunek w rolnictwie			
Miasta	26,062	3,427	3,998	28,980	0,073	62,540
Wieś	15,611	2,053	2,395	17,358	0,044	37,460
Suma	41,673	5,480	6,393	46,338	0,117	100,000

Ź r ó d ł o: opracowanie własne na podstawie tabl. 3.

Z kolei mnożąc — zgodnie ze wzorem (7) — wyznaczone prawdopodobieństwa hipotetyczne przez liczbę mieszkańców Polski, którzy w 2002 r. mieli własne źródło utrzymania (23073304), otrzymano liczebność hipotetyczną (tabl. 5).

TABL. 5. LICZEBNOŚĆ HIPOTETYCZNA (teoretyczna) W 2002 R.

Wyszczególnienie	Główne dochody					Suma
	z pracy			ze źródeł niezarobkowych	z własności	
	najemnej	na własny rachunek poza rolnictwem	na własny rachunek w rolnictwie			
Miasta	6013444	790672	922446	6686603	16930	14430094
Wieś	3601880	473589	552518	4005082	10140	8643210
Suma	9615324	1264261	1474964	10691685	27070	23073304

Ź r ó d ł o: opracowanie własne na podstawie tabl. 2 i 4.

Dysponując wszystkimi elementami macierzy liczebności empirycznej n_{ij} (tabl. 2) oraz macierzy liczebności teoretycznej n_{ij}^{*} (tabl. 5) można wyznaczyć poszczególne składniki statystyki testowej χ^2. Obliczona — na podstawie wzoru (8) — wartość statystyki χ^2 wyniosła 2395797. Czcionką pogrubioną zaznaczono największe wartości jej składników (tabl. 6).

TABL. 6. SKŁADNIKI STATYSTYKI TESTOWEJ CHI-KWADRAT W 2002 R.

Wyszczególnienie	Główne dochody					Suma
	z pracy			ze źródeł niezarobkowych	z własności	
	najemnej	na własny rachunek poza rolnictwem	na własny rachunek w rolnictwie			
Miasta	**97379**	**33114**	**765261**	1205	501	897461
Wieś	**162578**	**55285**	**1277626**	2012	836	1498337
Suma	259957	88399	2042887	3217	1337	2395798

Ź r ó d ł o: opracowanie własne na podstawie tabl. 2 i 5.

Mając obliczoną wartość χ^2 można przystąpić do weryfikacji hipotezy zerowej o niezależności cech będących przedmiotem analizy. Ponieważ $r=2$ oraz $s=5$, to liczba stopni swobody jest równa $(r-1)(s-1)=1\cdot 4=4$. Jeśli przyjmie się poziom istotności $\alpha=0{,}001$, to dla 4 stopni swobody odczytana z tablic rozkładu chi-kwadrat wartość krytyczna χ^2_{α} wynosiła 18,5. Porównując obliczoną wartość χ^2 z odpowiednią wartością krytyczną otrzymano $\chi^2=2395797>18{,}5=\chi^2_{\alpha}$.

Skoro uzyskano nierówność $\chi^2>\chi^2_{\alpha}$, to hipotezę H_0 o niezależności głównego źródła utrzymania mieszkańca Polski w 2002 r. od jego miejsca zamieszkania należy odrzucić na rzecz hipotezy alternatywnej, orzekając tym samym, że rozpatrywane cechy są zależne. Można zatem stwierdzić, że istnieje statystycznie istotny związek między częstotliwością występowania wymienionych źródeł dochodu a tym, czy osoba mieszka w mieście czy na wsi. Okazało się bowiem, że uzyskane odchylenia między liczebnością empiryczną i teoretyczną były wystarczająco duże, aby odrzucić przypuszczenie o niezależności. Należy się jednak liczyć z możliwością popełnienia błędu odrzucenia hipotezy H_0, pomimo tego że w rzeczywistości jest ona prawdziwa — prawdopodobieństwo popełnienia takiego błędu wynosi w tym przypadku 0,001. W teorii statystyki błąd ten nosi nazwę błędu I rodzaju.

Po przeprowadzeniu porównania poszczególnych prawdopodobieństw empirycznych znajdujących się w tabl. 3 z odpowiadającymi im prawdopodobieństwami teoretycznymi znajdującymi się w tabl. 4 można wyciągnąć wniosek, że

w 2002 r. osoby mieszkające w mieście względnie częściej czerpały dochody z pracy najemnej oraz z pracy na własny rachunek poza rolnictwem, natomiast osoby mieszkające na wsi relatywnie rzadziej. Z kolei osoby mieszkające na wsi relatywnie częściej osiągały dochody z pracy na własny rachunek w rolnictwie.

ANALIZA WPŁYWU MIEJSCA ZAMIESZKANIA NA ROZKŁAD GŁÓWNYCH ŹRÓDEŁ UTRZYMANIA LUDNOŚCI W POLSCE W 2011 R.

Drugim postawionym przed Autorkami zadaniem jest odpowiedź na pytanie, czy w 2011 r. istniała zależność między tym, czy osoba mieszka w mieście czy na wsi a tym, jakie jest jej główne źródło utrzymania. Ponownie weryfikacji będzie podlegała hipoteza H_0 stanowiąca, że badane dwie cechy są stochastycznie niezależne, wobec hipotezy alternatywnej H_1 orzekającej, że występuje stochastyczna zależność między tymi cechami. Dane niezbędne do przeprowadzenia weryfikacji za pomocą testu niezależności chi-kwadrat zebrano w tabl. 7[15].

TABL. 7. LICZEBNOŚĆ EMPIRYCZNA (rzeczywista) W 2011 R.

Wyszczególnienie	Główne dochody					Suma
	z pracy			ze źródeł niezarobkowych	z własności	
	najemnej	na własny rachunek poza rolnictwem	na własny rachunek w rolnictwie			
Miasta	7914764	1040088	97828	6117058	19500	15189238
Wieś	3986774	455084	1059658	3625777	16242	9143535
Suma	11901538	1495172	1157486	9742835	35742	24332773

Ź r ó d ł o: opracowanie własne na podstawie publikacji *Ludność…* (2013).

Na podstawie informacji zaprezentowanych w tabl. 7 obliczono poszczególne prawdopodobieństwa empiryczne. Następnie, zakładając stochastyczną niezależność cech, obliczono prawdopodobieństwa teoretyczne. Z kolei mnożąc wyznaczone prawdopodobieństwa hipotetyczne przez liczbę mieszkańców Polski w 2011 r. posiadających własne źródło utrzymania (24332773), otrzymano liczebność hipotetyczną. Mając wszystkie elementy macierzy liczebności empirycznej n_{ij} i macierzy liczebności teoretycznej n_{ij}^*, wyznaczono wartość statystyki testowej χ^2. Statystyka ta wyniosła w tym przypadku 1555791.

W tabl. 8—11 przedstawiono wyniki uzyskane ze wszystkich opisanych kroków obliczeniowych. W tabl. 11 czcionką pogrubioną zaznaczono największe składniki statystyki testowej χ^2.

[15] Zgodnie z danymi GUS w 2011 r. 24332773 mieszkańców Polski miało własne źródło utrzymania, a ich głównym źródłem utrzymania była: praca najemna, praca na własny rachunek w rolnictwie, praca na własny rachunek poza rolnictwem, własność albo źródło niezarobkowe.

TABL. 8. PRAWDOPODOBIEŃSTWA EMPIRYCZNE (rzeczywiste) W 2011 R. W %

Wyszczególnienie	Główne dochody					Suma
	z pracy			ze źródeł niezarobkowych	z własności	
	najemnej	na własny rachunek poza rolnictwem	na własny rachunek w rolnictwie			
Miasta	32,527	4,274	0,402	25,139	0,080	62,423
Wieś	16,384	1,870	4,355	14,901	0,067	37,577
Suma	48,911	6,144	4,757	40,040	0,147	100,000

Ź r ó d ł o: opracowanie własne na podstawie tabl. 7.

TABL. 9. PRAWDOPODOBIEŃSTWA HIPOTETYCZNE (teoretyczne) W 2011 R. W %

Wyszczególnienie	Główne dochody					Suma
	z pracy			ze źródeł niezarobkowych	z własności	
	najemnej	na własny rachunek poza rolnictwem	na własny rachunek w rolnictwie			
Miasta	30,532	3,836	2,969	24,994	0,092	62,423
Wieś	18,380	2,309	1,788	15,046	0,055	37,577
Suma	48,912	6,145	4,757	40,040	0,147	100,000

Ź r ó d ł o: opracowanie własne na podstawie tabl. 8.

TABL. 10. LICZEBNOŚĆ HIPOTETYCZNA (teoretyczna) W 2011 R.

Wyszczególnienie	Główne dochody					Suma
	z pracy			ze źródeł niezarobkowych	z własności	
	najemnej	na własny rachunek poza rolnictwem	na własny rachunek w rolnictwie			
Miasta	7429293	933331	722537	6081766	22311	15189238
Wieś	4472245	561841	434949	3661069	13431	9143535
Suma	11901538	1495172	1157486	9742835	35742	24332773

Ź r ó d ł o: opracowanie własne na podstawie tabl. 7 i 9.

TABL. 11. SKŁADNIKI STATYSTYKI TESTOWEJ CHI-KWADRAT W 2011 R.

Wyszczególnienie	Główne dochody					Suma
	z pracy			ze źródeł niezarobkowych	z własności	
	najemnej	na własny rachunek poza rolnictwem	na własny rachunek w rolnictwie			
Miasta	**31723**	**12211**	**540126**	205	354	584620
Wieś	**52699**	**20285**	**897258**	340	588	971171
Suma	84422	32496	1437384	545	942	1555791

Ź r ó d ł o: opracowanie własne na podstawie tabl. 7 i 10.

Skoro wyznaczono wartość statystyki χ^2, to można przystąpić do weryfikacji hipotezy zerowej o niezależności cech będących przedmiotem analizy. Wartość krytyczna χ^2_α odczytana z tablic rozkładu chi-kwadrat dla poziomu istotności α równego 0,001 i 4 stopni swobody wynosi 18,5. Porównując obliczoną wartość statystyki χ^2 z wartością krytyczną uzyskano $\chi^2 = 1555791 > 18{,}5 = \chi^2_\alpha$.

Spełniona została nierówność $\chi^2 > \chi^2_\alpha$, wobec czego hipotezę H_0 o niezależności rozkładu głównych źródeł utrzymania mieszkańców Polski w 2011 r. od miejsca zamieszkania należy odrzucić na rzecz hipotezy alternatywnej. Można więc stwierdzić, że istnieje statystycznie istotny związek między częstotliwością występowania wymienionych pięciu źródeł utrzymania a tym, czy dana osoba mieszka na wsi czy w mieście. Należy się jednak liczyć z możliwością popełnienia błędu odrzucenia hipotezy H_0, pomimo tego że w rzeczywistości jest ona prawdziwa — prawdopodobieństwo popełnienia takiego błędu wynosi 0,001. Jest to błąd I rodzaju.

Z porównania poszczególnych prawdopodobieństw empirycznych pokazanych w tabl. 8 z odpowiadającymi im prawdopodobieństwami teoretycznymi znajdującymi się w tabl. 9 można wyciągnąć wniosek, że w 2011 r. dla osób mieszkających w mieście praca najemna oraz praca na własny rachunek poza rolnictwem były stosunkowo częściej głównym źródłem utrzymania niż dla osób mieszkających na wsi. Z kolei dla osób mieszkających na wsi praca na własny rachunek w rolnictwie była stosunkowo częściej głównym źródłem utrzymania w porównaniu z tą grupą osób, które mieszkały w mieście.

PORÓWNANIE WPŁYWU MIEJSCA ZAMIESZKANIA NA ROZKŁAD ŹRÓDEŁ UTRZYMANIA LUDNOŚCI POLSKI W LATACH 2002 I 2011

We wcześniejszych rozdziałach opracowania udowodniono, że na zróżnicowanie Polaków pod względem występowania poszczególnych źródeł utrzymania w 2002 r. i w 2011 r. statystycznie istotny wpływ miało miejsce zamieszkania. Warto więc odpowiedzieć na pytanie, jak silna była ta zależność oraz czy zależność istniejąca w 2011 r. uległa w stosunku do 2002 r. zwiększeniu czy może zmniejszeniu. W tym celu należy określić siłę występujących zależności dla lat 2002 i 2011, a następnie dokonać stosownych porównań.

Do określenia siły zależności posłużył współczynnik *V* Cramera. Wartość tego współczynnika wyniosła w latach: 2002 — 0,32; 2011 — 0,25.

Można zatem stwierdzić, że miejsce zamieszkania w znacznym stopniu rzutowało na to, jakie było główne źródło utrzymania mieszkańca Polski tak w 2002 r., jak i w 2011 r. Należy jednak zauważyć, że fakt zamieszkiwania

przez daną osobę w mieście czy na wsi w mniejszym zakresie wpływał na zróżnicowanie mieszkańców Polski pod względem głównych źródeł utrzymania w 2011 r. niż było to w roku 2002. Dowodem na to jest niższa wartość współczynnika V Cramera obliczona dla 2011 r. od jego wartości obliczonej dla roku 2002.

W tabl. 12 pokazano różnice między prawdopodobieństwami empirycznymi i odpowiadającymi im prawdopodobieństwami teoretycznymi w 2002 r., a w tabl. 13 różnice między prawdopodobieństwami empirycznymi i teoretycznymi w 2011 r. Z kolei w tabl. 14 porównano poszczególne składniki statystyki testowej χ^2 otrzymane dla 2002 r. z analogicznymi wielkościami obliczonymi dla 2011 r.

TABL. 12. RÓŻNICE MIĘDZY PRAWDOPODOBIEŃSTWAMI EMPIRYCZNYMI I TEORETYCZNYMI W 2002 R. W P.PROC.

Wyszczególnienie	Główne dochody					Suma
	z pracy			ze źródeł niezarobkowych	z własności	
	najemnej	na własny rachunek poza rolnictwem	na własny rachunek w rolnictwie			
Miasta	3,317	0,701	−3,641	−0,389	0,013	0,000
Wieś	−3,317	−0,701	3,641	0,389	−0,013	0,000
Suma	0,000	0,000	0,000	0,000	0,000	0,000

Ź r ó d ł o: opracowanie własne na podstawie tabl. 3 i 4.

TABL. 13. RÓŻNICE MIĘDZY PRAWDOPODOBIEŃSTWAMI EMPIRYCZNYMI I TEORETYCZNYMI W 2011 R. W P.PROC.

Wyszczególnienie	Główne dochody					Suma
	z pracy			ze źródeł niezarobkowych	z własności	
	najemnej	na własny rachunek poza rolnictwem	na własny rachunek w rolnictwie			
Miasta	1,995	0,439	−2,567	0,145	−0,012	0,000
Wieś	−1,995	−0,439	2,567	−0,145	0,012	0,000
Suma	0,000	0,000	0,000	0,000	0,000	0,000

Ź r ó d ł o: opracowanie własne na podstawie tabl. 8 i 9.

TABL. 14. PORÓWNANIE SKŁADNIKÓW STATYSTYKI TESTOWEJ CHI-KWADRAT W LATACH 2002 I 2011

L a t a	Główne dochody					Suma
	z pracy			ze źródeł niezarobkowych	z własności	
	najemnej	na własny rachunek poza rolnictwem	na własny rachunek w rolnictwie			
Miasta						
2002	97379	33114	765261	1205	501	897461
2011	31723	12211	540126	205	354	584620

TABL. 14. PORÓWNANIE SKŁADNIKÓW STATYSTYKI TESTOWEJ CHI-KWADRAT W LATACH 2002 I 2011 (dok.)

Lata	Główne dochody					Suma
	z pracy			ze źródeł niezarobkowych	z własności	
	najemnej	na własny rachunek poza rolnictwem	na własny rachunek w rolnictwie			
	Wieś					
2002	162578	55285	1277626	2012	836	1498337
2011	52699	20285	897258	340	588	971171

Ź r ó d ł o: opracowanie własne na podstawie tabl. 6 i 11.

Porównanie wyników znajdujących się w tabl. 12 i 13 pozwala na wyciągnięcie wniosku, że w przypadku każdego z prawdopodobieństw empirycznych miało ono wartość bliższą prawdopodobieństwu teoretycznemu w 2011 r. niż było to w roku 2002. Z kolei na podstawie tabl. 14 można stwierdzić, że wszystkie składniki statystyki χ^2 obliczonej dla 2011 r. były dużo niższe od odpowiadających im składników statystyki χ^2 odnoszącej się do roku 2002. Nie ulega więc wątpliwości, że na przestrzeni dziewięciu lat różnice między mieszkańcami miast i wsi w Polsce w zakresie tego, czy dochody ich pochodzą z pracy najemnej, z pracy na własny rachunek poza rolnictwem, z pracy na własny rachunek w rolnictwie, ze źródeł niezarobkowych czy z własności uległy znacznemu zatarciu.

Podsumowanie

Głównym celem artykułu jest udzielenie odpowiedzi na pytanie, jak zmienił się wpływ miejsca zamieszkania na rodzaj głównego źródła utrzymania mieszkańców Polski między latami 2002 i 2011. Przeprowadzone badanie pozwoliło sformułować następujące wnioski:

- w 2002 r. i w 2011 r. istniała w Polsce statystycznie istotna zależność między tym, czy osoba mieszka w mieście czy na wsi a tym, jakie jest jej główne źródło utrzymania. Pobieranie dochodów z pracy najemnej bądź z pracy na własny rachunek poza rolnictwem było typowe dla mieszkańców miast, natomiast pobieranie dochodów z pracy na własny rachunek w rolnictwie było charakterystyczne dla mieszkańców wsi;
- wpływ miejsca zamieszkania na rozkład głównych źródeł utrzymania ludności Polski w roku 2011 był zdecydowanie mniejszy niż w roku 2002;
- różnica między udziałem ludności miast, dla której praca najemna była głównym źródłem utrzymania a udziałem ludności wsi, dla której praca najemna była głównym źródłem utrzymania uległa na przestrzeni dziewięciu lat znacznej redukcji. Prawidłowość taką zaobserwowano nie tylko w przy-

padku pracy najemnej, ale również pracy na własny rachunek poza rolnictwem i na własny rachunek w rolnictwie, własności oraz źródeł niezarobkowych.

W toku przeprowadzonych badań udowodniono, że w 2011 r. struktury ludności w mieście i na wsi — pod względem głównych źródeł dochodów — mniej się od siebie różniły niż można było to zaobserwować w 2002 r. Hipotezę badawczą mówiącą, że fakt mieszkania w mieście albo na wsi w coraz mniejszym zakresie determinuje rozkład mieszkańców Polski pod względem rodzajów posiadanych źródeł utrzymania zweryfikowano pozytywnie. Pozwoliło to na stwierdzenie, że polska wieś i polskie miasta na przestrzeni rozpatrywanych dziewięciu lat uległy pod względem rozpatrywanej cechy znacznemu ujednoliceniu.

dr Anna Turczak — *Zachodniopomorska Szkoła Biznesu w Szczecinie*, **dr Patrycja Zwiech** — *Uniwersytet Szczeciński*

LITERATURA

Aczel A. D. (2000), *Statystyka w zarządzaniu*, PWN, Warszawa

Buga J., Kassyk-Rokicka H. (2008), *Podstawy statystyki opisowej*, Wyższa Szkoła Finansów i Zarządzania w Warszawie

Jóźwiak J., Podgórski J. (1995), *Statystyka od podstaw*, Polskie Wydawnictwo Ekonomiczne, Warszawa

Kot S., Jakubowski J., Sokołowski A. (2007), *Statystyka. Podręcznik dla studiów ekonomicznych*, Centrum Doradztwa i Informacji „Difin", Warszawa

Krysicki W., Bartos J., Dyczka W., Królikowska K., Wasilewski M. (2003), *Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach. Część II*, PWN, Warszawa

Kukuła K. (2003), *Elementy statystyki w zadaniach*, PWN, Warszawa

Ludność i gospodarstwa domowe. Stan i struktura społeczno-ekonomiczna 2002 (2003), GUS

Ludność i gospodarstwa domowe. Stan i struktura społeczno-ekonomiczna 2011 (2013), GUS

Ostasiewicz S., Rusnak Z., Siedlecka U. (1995), *Statystyka. Elementy teorii i zadania*, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu

Podgórski J. (2005), *Statystyka dla studiów licencjackich*, PWE, Warszawa

Pułaska-Turyna B. (2005), *Statystyka dla ekonomistów*, Centrum Doradztwa i Informacji „Difin", Warszawa

Zeliaś A. (2000), *Metody statystyczne*, PWE, Warszawa

SUMMARY

The aim of the study is to examine the strength of the association between the place of residence of the Polish population and their main source of income on the basis of the last two censuses carried out in 2002 and 2011. The study used Pearson chi-square test of independence and Cramer's V coefficient to assess

the relationship between the changes in residence place and the main source of income, which occurred between the two last censuses period. The analysis showed that the difference between urban and rural areas, depending on the source of income has been over nine years significantly reduced. This regularity was observed by comparisons between the study years the shares of the population, which wage labor was the main source of income, as well as the shares of the population, the main source of income was a self-employed outside agriculture, self-employment in agriculture, property or non-commercial sources.

РЕЗЮМЕ

Целью статьи является обследование прочности связи между местом проживания населения Польши и их главным источником средств на содержание на основе двух последних всеобщих переписей населения проводимых в 2002 г. и в 2011 г. В обследовании был использован критерий независимости хи-квадрат Пирсона и коэффициент V Крамэра для оценки изменения зависимости между местом проживания и главным источником доходов, которая имела место в межпереписный период.

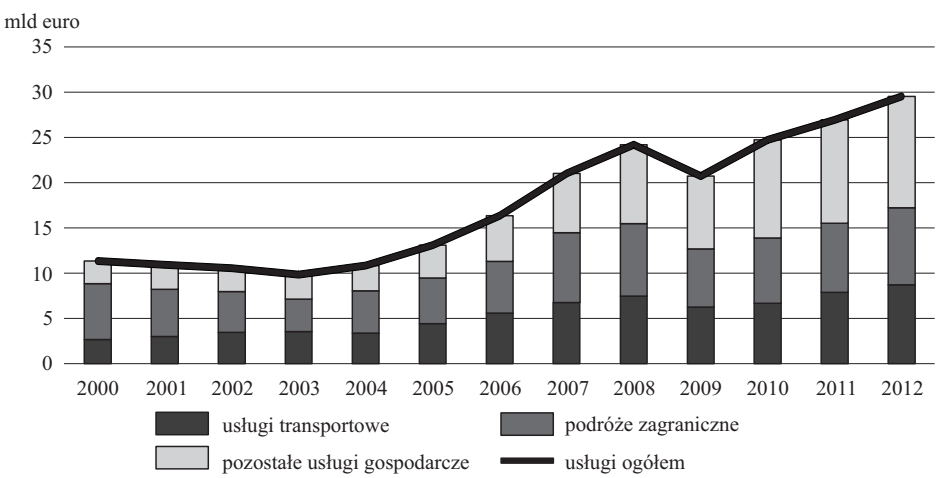
Проведенный анализ показал, что разница между городом и деревней в зависимости от источников дохода за 9 лет значительно уменьшилась. Эта точность наблюдается при сопоставлении в обследуемых годах доли населения, которого наемный труд был главным источником средств на содержание, а также доли населения, которого главным источником средств на содержание является самозанятость кроме сельского хозяйства, самозанятость в сельском хозяйстве, частная собственность или работа без зарплаты.

Łukasz MATUSZCZAK

Konkurencyjność polskiego eksportu usług

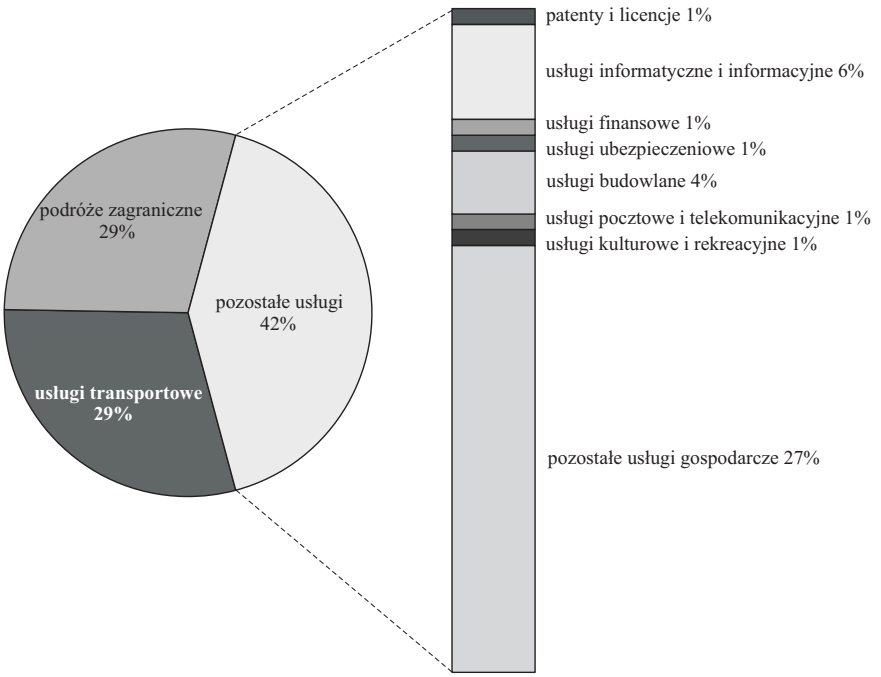
Celem opracowania jest zidentyfikowanie grup usługowych, w których Polska wykazuje ujawnioną przewagą komparatywną (*Manual*..., 2011). Wagę podjętej w opracowaniu tematyki podkreślają publikowane podstawowe dane makroekonomiczne dotyczące polskiego eksportu usług. Udział eksportu usług w odniesieniu do PKB na początku badanego okresu (2000 r.) stanowił 6,09%, a na końcu tego okresu (2012 r.) — 7,7%. W 2000 r. eksport usług naszej gospodarki wynosił 11320 mln euro, natomiast w 2012 r. notowano prawie trzykrotny jego wzrost — wartość ta osiągnęła 29517 mln euro. Na początku analizowanego okresu dominował eksport usług związanych z podróżami zagranicznymi osiągając 55% ogółu przychodów; usługi transportowe, które stanowiły 23% i pozostałe usługi — 22%. Proporcje te zachowały się w przybliżeniu aż do 2007 r., w kolejnym roku dominować zaczęły pozostałe usługi, przy czym podróże zagraniczne traciły na znaczeniu na rzecz stopniowego przyrostu udziału usług transportowych. Tendencja ta utrzymywała się do roku 2012. Rozwój międzynarodowego handlu polskimi usługami obrazuje wykr. 1. W strukturze polskiego eksportu usług dominowały transport i podróże zagraniczne (po 29%), bliską im wartość osiągnęły również pozostałe usługi gospodarcze (27%), natomiast usługi informatyczne i informacyjne stanowiły tylko 6% (wykr. 2).

Wykr. 1. PRZYCHODY Z TYTUŁU EKSPORTU USŁUG



Ź r ó d ł o: opracowanie własne.

Wykr. 2. STRUKTURA POLSKIEGO EKSPORTU USŁUG W 2012 R.



Ź r ó d ł o: jak przy wykr. 1.

Uzyskane w badaniu wyniki porównano z miarami uzyskanymi dla wybranych krajów sąsiadujących z Polską (Niemiec, Czech, Słowacji). Dodatkowo postanowiono sprawdzić, jak kształtował się eksport poszczególnych kategorii usługowych według województw. Ostatnim z postawionych celów było określenie, czy dany region wyróżnia się w eksporcie usług, ogółem, transportowych i pozostałych.

MIARA PRZEWAGI KOMPARATYWNEJ W EKSPORCIE USŁUG[1]

Zbadano, w jakich dziedzinach międzynarodowego handlu usługami specjalizuje się Polska. W tym celu użyto wskaźnika ujawnionej przewagi komparatywnej (*revealed comparative advantage — RCA*). Miarę tę wybrano ze względu na jej popularność w literaturze ekonomicznej. Formuła oparta jest na teorii przewagi komparatywnej stworzonej i spopularyzowanej przez węgierskiego ekonomistę B. Ballassę w 1958 r. *RCA* wyraża względną przewagę kraju w międzynarodowym eksporcie danego dobra/usługi w porównaniu do całkowitego udziału tego kraju w eksporcie grupy referencyjnej. W tym przypadku jako grupę odniesienia przyjęto kraje strefy euro (wybrane ze względu na największy udział w polskiej

[1] Balassa (1986).

wymianie z tytułu usług). Formuła ujawnionej przewagi komparatywnej jest następująca:

$$RCA_j^A = \frac{\frac{ex_j^A}{ex^A}}{\frac{ex_j^{REF}}{ex^{REF}}}$$

gdzie:

RCA_j^A — wskaźnik ujawnionej przewagi komparatywnej kraju A w komponencie j,
ex_j^A — eksport kraju A w komponencie j,
ex^A — całkowity eksport usług kraju A,
ex_j^{REF} — eksport usług grupy krajów będących punktem odniesienia w komponencie j,
ex^{REF} — całkowity eksport usług grupy krajów będących punktem odniesienia.

Kraj wykazuje ujawnioną przewagę komparatywną, gdy $RCA>1$. Oznacza to, że analizowane państwo specjalizuje się w eksporcie danej usługi względem grupy odniesienia. Jeśli $RCA\leq1$, to kraj A jest względnie niewyspecjalizowany w eksporcie danej usługi.

W opracowaniu wykorzystano dane NBP, pochodzące z bilansu płatniczego oraz informacje o grupie krajów odniesienia zaczerpnięte z bazy Eurostatu. Te ostatnie dotyczyły danych uwzględniających 19 krajów europejskich (strefa euro oraz Polska i Czechy). Zbiory obejmowały lata 2000–2011 i zawierały dane o przychodach z międzynarodowego handlu usługami, a dla 2012 r. uwzględniono dane NBP dla Polski oraz wstępne dane Eurostatu dla krajów odniesienia.

PRZEWAGA KOMPARATYWNA POLSKI W EKSPORCIE USŁUG

Na początku badanego okresu (2000 r.) Polska prezentowała ujawnioną przewagę komparatywną w trzech kategoriach usług: pocztowych i kurierskich, transportowych oraz w podróżach zagranicznych. W kolejnych dwóch latach polskie usługi pocztowe i kurierskie straciły pozycję konkurencyjną, z kolei na znaczeniu zyskały usługi budowlane oraz umocniły pozycję konkurencyjną usługi transportowe i podróże zagraniczne (Hagemejer, 2006).

W tabl. 1 zaprezentowano wskaźniki *RCA* dla lat 2000—2012. Przyjęto podział usług na grupy zgodny z wytycznymi Międzynarodowego Funduszu Walutowego (*Balance…*, 2009). Tłustym drukiem wyróżniono obserwacje o $RCA>1$.

TABL. 1. WSKAŹNIKI *RCA* DLA POLSKI

Usługi	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Pocztowe i telekomunikacyjne	**1,28**	0,95	0,73	0,93	**1,02**	0,85	0,75	0,70	0,64	0,76	0,56	0,53	0,44
Budowlane	0,94	**1,39**	**1,85**	**2,35**	**1,95**	**1,88**	**1,98**	**1,89**	**1,76**	**1,69**	**1,52**	**1,52**	**1,48**
Finansowe	0,36	0,77	0,28	0,31	0,25	0,25	0,16	0,18	0,23	0,23	0,25	0,18	0,18
Informatyczne i informacyjne	0,17	0,19	0,21	0,23	0,29	0,24	0,35	0,39	0,40	0,43	0,64	0,73	0,79
Patenty i licencje	0,14	0,23	0,15	0,08	0,07	0,11	0,06	0,10	0,16	0,07	0,13	0,20	0,15
Pozostałe usługi gospodarcze	0,52	0,50	0,48	0,54	0,62	0,68	0,70	0,72	0,87	0,95	**1,11**	0,99	0,89
Kulturalne i rekreacyjne	0,47	0,69	0,62	0,57	0,80	0,71	0,90	0,86	0,84	0,62	**1,12**	**1,12**	0,85
Rządowe	0,01	0,05	0,02	0,02	0,13	0,17	0,18	0,21	0,32	0,34	0,03	0,02	0,00
Transportowe	0,01	0,05	0,02	0,02	0,13	0,17	0,18	0,21	0,32	0,34	0,03	0,02	0,00
Podróże zagraniczne	**1,55**	**1,42**	**1,41**	**1,23**	**1,49**	**1,41**	**1,33**	**1,47**	**1,39**	**1,55**	**1,42**	**1,41**	**1,23**

Ź r ó d ł o: opracowanie własne.

W 2004 r. notowano (w roku przystąpienia Polski do Unii Europejskiej (UE)) pogorszenie się konkurencyjności naszego kraju w porównaniu do 2003 r. na rynku usług budowlanych (niemniej jednak wskaźnik *RCA* nadal znajdował się powyżej jedności). Jak pokazuje tabl. 1, to raczej rok 2003 charakteryzował się dużą wartością *RCA* w stosunku do reszty przedstawionego szeregu, zatem spadek konkurencyjności w 2004 r. nie był niczym szczególnym. Sytuacja spadkowa (w 2004 r.) miała miejsce również w usługach finansowych, transporcie oraz podróżach zagranicznych. W drugim roku po przystąpieniu Polski do UE sytuacja w wymianie handlowej usług zaliczanych do pracochłonnych, tj. budowlanych i transportowych, poprawiła się. W obu kategoriach *RCA* uległo podwyższeniu.

Kryzys finansowy, którego początek miał miejsce w 2008 r., nie wpłynął w znaczący sposób na konkurencyjność polskich usług na arenie międzynarodowej. Prawie połowa wyszczególnionych kategorii poprawiła konkurencyjność (wynikało to prawdopodobnie z pogorszenia się konkurencyjności cenowej innych krajów wskutek deprecjacji złotego), spadki zaś w konkurencyjności były niewielkie.

Od początku prezentowanego okresu, aż do 2010 r., *RCA* usług kulturalnych i rekreacyjnych oraz pozostałych usług gospodarczych wykazywał systematyczną tendencję rosnącą. W 2010 r. oba wskaźniki były większe od jedności, co wskazywało na ujawnienie przewagi komparatywnej w eksporcie obu kategorii. Tendencja ta nie utrzymała się długo, w przypadku pozostałych usług gospodarczych *RCA* spadło poniżej jedności już w następnym roku. Podobna sytuacja miała miejsce w usługach kulturalnych i rekreacyjnych (spadek *RCA* poniżej jedności miał miejsce w 2011 r.). Eksport obu kategorii sprawiał, że wartości *RCA* wprawdzie nie przekroczyły wartości granicznej, ale utrzymywały się w jej pobliżu (pozwala to mieć nadzieję na uzyskanie w następnych latach pozycji konkurencyjnej w eksporcie omawianych kategorii).

PORÓWNANIE KONKURENCYJNOŚCI POLSKIEGO EKSPORTU USŁUG Z KRAJAMI SĄSIEDZKIMI

Do analizy porównawczej, jak wspomniano wcześniej, przyjęto Czechy, Słowację oraz Niemcy. Dwa pierwsze kraje wybrano na podstawie podobieństwa z gospodarką polską (wspólna historia postsocjalistyczna) oraz wynikające z sąsiedztwa. Gospodarka niemiecka posłużyła w analizie jako kraj referencyjny (najsilniejsza gospodarka strefy euro). W pierwszej grupie usług zaprezentowano te kategorie, w których Polska uzyskała ujawnioną przewagę komparatywną, były to usługi budowlane, transportowe oraz podróże zagraniczne.

Jako pierwsze omówiono usługi budowlane. Z dokonanych obliczeń (tabl. 2) wynika, że w 2000 r. wszystkie kraje miały szansę na uzyskanie bardzo dobrej pozycji konkurencyjnej w handlu usługami budowlanymi (*RCA* bliskie jedności). Do grona krajów specjalizujących się w tych usługach w kolejnym roku dołączyła jedynie Polska; w roku 2010 konkurencyjną pozycję osiągnęły wszystkie analizowane kraje. Polskie usługi w tej kategorii od 2006 r. notowały sukcesywny spadek wskaźnika *RCA* (jednak nie na tyle duży, aby wartość ta spadła poniżej jedności). Podobna sytuacja dotyczyła Niemiec, jednak w tym przypadku spadek okazał się większy. Zarówno Czechy, jak i Słowacja uzyskały lepszą pozycję konkurencyjną niż rok wcześniej.

TABL. 2. WARTOŚCI *RCA* W EKSPORCIE USŁUG BUDOWLANYCH WEDŁUG KRAJÓW

Kraje	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Czechy	0,81	0,71	0,59	0,52	0,48	0,71	0,52	0,63	0,73	0,82	**1,75**	**1,40**	**1,43**
Niemcy	**1,60**	**1,51**	**1,66**	**2,06**	**1,91**	**2,24**	**2,19**	**2,05**	**2,05**	**1,91**	**1,84**	**1,76**	**1,48**
Polska	0,94	**1,39**	**1,85**	**2,35**	**1,95**	**1,88**	**1,98**	**1,89**	**1,76**	**1,69**	**1,52**	**1,52**	**1,48**
Słowacja	0,97	0,77	0,61	0,92	**1,20**	0,95	0,53	0,58	0,70	0,66	**1,05**	**1,33**	**1,69**

Źródło: jak przy tabl. 1.

Kolejną kategorią były usługi transportowe. W latach 2000 i 2001 tylko Czechy nie wykazywały ujawnionej przewagi komparatywnej w międzynarodowym handlu usługami transportowymi. Kolejne dwa lata pozwoliły wszystkim analizowanym krajom specjalizować się w świadczeniu usług transportowych. W 2004 r. omawiane kraje — z wyjątkiem Niemiec — odnotowały spadek konkurencyjności, przy czym nie był on na tyle duży, aby utracić ujawnioną przewagę komparatywną (Mongiało, 2007). Ten pozytywny stan rzeczy mógł wynikać m.in. z tranzytowego położenia geograficznego analizowanych krajów (między Rosją i UE). W 2012 r. Polska zajęła pierwsze miejsce wśród omawianych państw w tej kategorii.

TABL. 3. WARTOŚCI *RCA* W EKSPORCIE USŁUG TRANSPORTOWYCH WEDŁUG KRAJÓW

Kraje	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Czechy	0,88	0,99	**1,18**	**1,42**	**1,35**	**1,17**	**1,14**	**1,21**	**1,12**	**1,24**	**1,17**	**1,16**	**1,19**
Niemcy	1,00	**1,07**	**1,10**	**1,13**	**1,13**	**1,19**	**1,14**	**1,14**	**1,15**	**1,15**	**1,17**	**1,10**	**1,09**
Polska	**1,02**	**1,26**	**1,56**	**1,83**	**1,49**	**1,59**	**1,60**	**1,50**	**1,40**	**1,55**	**1,29**	**1,40**	**1,50**
Słowacja	**1,88**	**1,66**	**1,99**	**2,19**	**1,91**	**1,72**	**1,65**	**1,49**	**1,56**	**1,53**	**1,48**	**1,54**	**1,42**

Źródło: jak przy tabl. 1.

Ostatnią kategorią były podróże zagraniczne. Od początku analizowanego okresu *RCA*>1 notowały Czechy i Polska. Specjalizacja naszej gospodarki w świadczeniu usług z tytułu podróży zagranicznych miała związek ze stosunkowo dużym udziałem podróży jednodniowych osób dokonujących zakupów w Polsce. Od 2000 r. (po trwającym 4 lata pogarszaniu się konkurencyjności) do grupy krajów specjalizujących się w usługach z tytułu podróży zagranicznych zaliczono powtórnie Słowację (Mongiało, 2007). Na atrakcyjność usług świadczonych przez kraje środkowoeuropejskie nie miał wpływu światowy kryzys finansowy.

TABL. 4. WARTOŚCI *RCA* W EKSPORCIE USŁUG ZWIĄZANYCH Z PODRÓŻAMI ZAGRANICZNYMI WEDŁUG KRAJÓW

Kraje	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Czechy	**1,24**	**1,33**	**1,38**	**1,55**	**1,49**	**1,49**	**1,58**	**1,60**	**1,50**	**1,52**	**1,52**	**1,42**	**1,47**
Niemcy	0,60	0,61	0,61	0,63	0,65	0,65	0,66	0,65	0,66	0,63	0,65	0,63	0,65
Polska	**1,55**	**1,42**	**1,41**	**1,23**	**1,49**	**1,41**	**1,33**	**1,47**	**1,39**	**1,31**	**1,30**	**1,24**	**1,34**
Słowacja	0,59	**1,01**	0,85	0,89	0,83	1,00	**1,06**	**1,15**	**1,28**	**1,56**	**1,71**	**1,64**	**1,51**

Źródło: jak przy tabl. 1.

Kolejną wyszczególnioną grupą usług były usługi mające największą perspektywę rozwoju, do których zaliczono: pozostałe usługi gospodarcze, usługi kulturalne i rekreacyjne oraz usługi informatyczne i informacyjne. Grupę tę wyselekcjonowano na podstawie zaobserwowanego ciągłego trendu wzrostowego w wyliczonym mierniku *RCA*.

Jako pierwszą opisano sytuację w pozostałych usługach gospodarczych. W tej kategorii od początku analizowanego okresu do 2005 r. Polska plasowała się na ostatnim miejscu wśród porównywanej grupy krajów, jednak w ciągu omawianych lat konkurencyjność naszego kraju w tym zakresie sukcesywnie zyskiwała na znaczeniu. Pokrywa się to z rosnącym udziałem pozostałych usług gospodarczych w całości polskiej międzynarodowej wymiany usług. Na rozwój usług świadczonych przez Polskę w tej kategorii również nie miał wpływu światowy kryzys finansowy. W badanym okresie tylko Niemcy wykazywały względną specjalizację w handlu z tytułu pozostałych usług gospodarczych. Polskie usługi z omawianej kategorii były konkurencyjne w 2010 r., zaś w 2011 r. wskaźnik *RCA* był bliski jedności. Patrząc na historię kształtowania się *RCA* (tabl. 5) można mieć nadzieję, że usługi te w najbliższym czasie zaczną stawać się coraz bardziej konkurencyjne.

TABL. 5. WARTOŚCI *RCA* W EKSPORCIE POZOSTAŁYCH USŁUG GOSPODARCZYCH WEDŁUG KRAJÓW

Kraje	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Czechy	0,84	0,81	0,82	0,72	0,71	0,75	0,70	0,75	0,97	0,91	0,95	0,96	0,88
Niemcy	**1,31**	**1,16**	**1,02**	**1,06**	**1,29**	**1,22**	**1,15**	**1,17**	**1,18**	**1,16**	**1,18**	**1,23**	**1,16**
Polska	0,52	0,50	0,48	0,54	0,62	0,68	0,70	0,72	0,87	0,95	**1,11**	0,99	0,89
Słowacja	0,85	0,65	0,71	0,53	0,73	0,73	0,65	0,64	0,64	0,47	0,54	0,49	0,75

Źródło: jak przy tabl. 1.

Ze względu na dużą różnorodność świadczonych usług omawianej grupy, przedstawiono bardziej szczegółową analizę pozostałych usług. Głównymi podgrupami usługowymi są: merchanting, leasing operacyjny, usługi świadczone przez profesjonalistów i usługi techniczne. W skład ostatniej wymienionej podgrupy wchodzą m.in.: usługi prawne, księgowe, public relations, reklama i badania rynku, badania i rozwój, usługi architektoniczne, inżynieryjne i techniczne, rolnicze oraz wydobywcze i usługi pomiędzy przedsiębiorstwami powiązanymi (*Manual…*, 2011).

Jak pokazuje tabl. 5, w 2009 r. Polska była bliska uzyskania wskaźnika *RCA* na poziomie większym od jedności i osiągnęła to w 2010 r. W następnym roku wskaźnik ujawnionej przewagi komparatywnej spadł, ale utrzymywał się na poziomie zbliżonym do jedności. Największy wpływ na kształtowanie się polskiej pozycji konkurencyjnej na rynku pozostałych usług gospodarczych miały (od 2009 r.) usługi prawne oraz związane z księgowością. Od roku 2010 do tej grupy dołączyły usługi dotyczące reklamy i badania opinii publicznej. Jednak ten nagły skok w konkurencyjności niektórych usług mógł wynikać z niedostatecznie szczegółowych informacji dla naszej gospodarki oraz dla krajów odniesienia w bazie Eurostatu; niektóre informacje dostępne były od roku 2009. Od tego właśnie okresu Polska notowała ujawnioną przewagę komparatywną w usługach prawnych oraz księgowych, audytu i konsultacji podatkowych. Można to tłumaczyć tym, że te rodzaje usług są często outsorcowane przez przedsiębiorstwa (w tym z zagranicy — przy wykonywaniu tych usług nie występuje bariera odległości, gdyż rezultaty pracy mogą być przekazane przez Internet). Mimo rosnącej specjalizacji w niektórych usługach spadek w 2011 r. *RCA* dla pozostałych usług gospodarczych był spowodowany faktem, że największy wartościowy udział eksportu w kategorii pozostałych usług gospodarczych miały takie usługi, jak: merchanting i pozostałe usługi związane z handlem, usługi architektoniczne i inżynieryjne oraz usługi pomiędzy przedsiębiorstwami powiązanymi. W wyniku stosowanego sposobu kalkulacji merchantingu w okresie 2001—2004 osiągnięto wartość $RCA<0$, zatem interpretacja tego wyniku nie jest możliwa. Ze względu na brak szczegółowych danych w kontekście krajów odniesienia zrezygnowano z prezentacji wartości wskaźników *RCA* dla roku 2012.

TABL. 6. WARTOŚCI *RCA* W EKSPORCIE WEDŁUG POZOSTAŁYCH USŁUG GOSPODARCZYCH

Usługi	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Pozostałe usługi gospodarcze	0,52	0,50	0,48	0,54	0,62	0,68	0,70	0,72	0,87	0,95	**1,11**	0,99
w tym:												
merchanting, usługi związane z handlem	.	–0,11	–0,04	–0,07	–0,06	0,20	0,06	0,16	0,39	0,11	0,15	0,12
leasing operacyjny	0,02	0,30	0,22	0,17	0,29	0,12	0,17	0,15	0,15	0,05	0,11	0,13
usługi prawne	.	.	.	.	.	.	.	.	.	**2,43**	**3,66**	**4,05**
usługi księgowe, audyt, konsultacje podatkowe	.	.	.	.	.	.	.	.	.	**2,26**	**2,09**	**2,29**

TABL. 6. WARTOŚCI *RCA* W EKSPORCIE WEDŁUG POZOSTAŁYCH USŁUG GOSPODARCZYCH (dok.)

Usługi	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
doradztwo gospodarcze oraz badania opinii publicznej	.	.	.	.	.	.	.	.	.	0,25	0,28	0,46
reklama i badanie opinii publicznej	0,29	0,28	0,50	0,49	0,34	0,44	0,69	0,24	0,48	0,86	1,27	1,99
badania i rozwój	0,48	0,07	0,08	0,10	0,17	0,13	0,28	0,23	0,11	0,26	0,34	0,33
usługi architektoniczne, inżynierskie i inne techniczne	0,00	0,56	0,62	0,53	0,49	0,52	0,59	0,71	0,46	0,60	0,78	0,92
rolnicze, górnicze oraz przetwarzania na miejscu	0,00	**4,52**	**1,14**	**2,67**	**2,67**	**2,54**	**10,3**	**3,40**	0,41	0,71	0,83	**1,35**
pozostałe usługi gospodarcze	**2,63**	**2,25**	**3,82**	**3,28**	**3,11**	**2,93**	**3,81**	**2,10**	**1,25**	0,84	0,96	**1,34**
usługi pomiędzy przedsiębiorstwami powiązanymi	0,00	0,00	0,00	0,00	0,01	0,03	0,04	0,04	0,10	0,47	0,31	0,36

U w a g a. Kropka oznacza brak danych.
Ź r ó d ł o: jak przy tabl. 1.

Kolejną kategorią usług zaliczonych do grupy usług o największej perspektywie rozwoju były usługi kulturalne i rekreacyjne. W tej kategorii Polska przez pierwsze 6 lat notowała powolny, ale sukcesywny rozwój konkurencyjności świadczonych usług. Wśród krajów Europy Środkowej prym wiodła Słowacja. W 2010 r. do grona krajów o wskaźniku *RCA* większym od jedności dołączyły Czechy (powtórnie) i Polska. W ostatnim roku prezentowanego okresu nasza gospodarka straciła pozycję konkurencyjną w świadczeniu omawianych usług. Podobna sytuacja miała miejsce w pozostałych analizowanych krajach.

Kształtowanie się wartości *RCA* na przestrzeni pokazywanych w artykule lat wskazuje, że obecnie obserwowany jest trend wzrostowy. Nie jest to wprawdzie tendencja sukcesywnych wzrostów (z okresu na okres), ale charakteryzuje się naprzemiennymi spadkami i wzrostami. W związku z tym można mieć nadzieję, że nasza gospodarka w najbliższym czasie znów uzyska ujawnioną przewagę komparatywną.

TABL. 7. WARTOŚCI *RCA* W EKSPORCIE USŁUG KULTURALNYCH I REKREACYJNYCH WEDŁUG KRAJÓW

K r a j e	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Czechy	**2,70**	**2,28**	**2,43**	**1,50**	**2,47**	0,94	**1,04**	**1,45**	0,81	0,89	**1,10**	**1,06**	0,93
Niemcy	0,46	0,58	0,53	0,85	0,78	0,89	0,62	0,65	0,62	0,66	0,54	0,33	0,29
P o l s k a	0,47	0,69	0,62	0,57	0,80	0,71	0,90	0,86	0,84	0,62	**1,12**	**1,12**	0,85
Słowacja	**2,31**	**2,41**	**2,37**	**2,46**	**3,92**	**3,45**	**3,55**	**5,89**	**2,14**	**1,54**	**1,43**	**1,98**	0,76

Ź r ó d ł o: jak przy tabl. 1.

Ze względu na różnorodność usług występujących w tej kategorii, zaprezentowano bardziej szczegółową jej analizę. W usługach kulturalnych i rekreacyjnych obowiązuje podział na dwie grupy — usługi audiowizualne i pozostałe

usługi. Ostatnią grupę można podzielić bardziej szczegółowo na usługi edukacyjne, zdrowotne oraz pozostałe. W latach 2000—2008 Polska, jak również inne kraje europejskie, nie raportowały do Eurostatu danych o eksporcie pozostałych usług kulturalnych i rekreacyjnych. Po rozpoczęciu przekazywania tych informacji w 2009 r. polskie usługi zyskały w tej dziedzinie ujawnioną przewagę komparatywną, podobna sytuacja miała miejsce zarówno w Czechach i na Słowacji. Niemcy nie zajmowały się w istotnym stopniu eksportem usług kulturalnych i rekreacyjnych. Jak pokazuje tabl. 7, w roku 2012 Polska straciła ujawnioną przewagę komparatywną w eksporcie usług kulturalnych i rekreacyjnych. Z powodu braku niektórych danych za 2012 r. nie można stwierdzić, które komponenty mogłyby poprawić konkurencyjność całej kategorii. Nie ulega wątpliwości, że ta kategoria ma szansę w przyszłości na powtórne uzyskanie *RCA*>1, wynika to z sukcesywnego wzrostu wskaźnika *RCA* na przestrzeni prezentowanego okresu.

TABL. 8. WARTOŚCI *RCA* W EKSPORCIE WEDŁUG USŁUG KULTURALNYCH I REKREACYJNYCH

Usługi	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Kulturalne i rekreacyjne	0,47	0,69	0,62	0,57	0,80	0,71	0,90	0,86	0,84	0,62	**1,12**	**1,12**
w tym:												
audiowizualne i związane z audiowizualnymi	0,00	0,34	0,68	**1,54**	**1,07**	0,40	0,37	0,27	0,27	0,22	0,27	0,44
edukacyjne	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	**7,23**	**4,46**	**4,36**
zdrowotne	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	**2,92**	**2,75**	**6,08**
pozostałe	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	**2,51**	**2,00**	**1,98**

Ź r ó d ł o: jak przy tabl. 1.

Ostatnią kategorią zaliczaną do grupy usług o największej perspektywie rozwoju były usługi informatyczne i informacyjne. W przypadku Polski od 2000 r. obserwowano systematyczny wzrost wskaźnika *RCA*, jednak był on mniejszy od jedności i dlatego nie uzyskaliśmy pozycji konkurencyjnej. Do 2004 r. tylko Niemcy wykazywały ujawnioną przewagę komparatywną w odniesieniu do strefy euro. W latach 2005 i 2006 przewagę komparatywną uzyskały Czechy (znaczący wzrost wartości *RCA* z 0,29 w 2004 r. do 1,19 w 2006 r.). W ostatnim badanym roku tylko Czechy świadczyły usługi informatyczne i informacyjne na dostatecznie konkurencyjnych warunkach. Można domniemywać, że w ciągu kilku lat Polska będzie w stanie osiągnąć dostatecznie konkurencyjną pozycje rynkową, wartość *RCA* wzrosła bowiem o ponad 100% na przestrzeni lat 2007—2012. W ostatnich latach omawianego okresu bardzo blisko wykazania ujawnionej przewagi komparatywnej były Niemcy oraz Słowacja, z kolei Polska prezentowała się najsłabiej. Jednak obserwowany ciągły wzrost wartości prezentowanego wskaźnika klasyfikuje ten rodzaj usług na trzecim miejscu wśród usług o największym potencjale rozwoju.

TABL. 9. WARTOŚCI *RCA* W EKSPORCIE USŁUG INFORMATYCZNYCH I INFORMACYJNYCH WEDŁUG KRAJÓW

Kraje	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Czechy	0,41	0,39	0,44	0,19	0,29	**1,10**	**1,19**	0,95	0,96	0,95	0,81	1,00	**1,13**
Niemcy	**1,28**	**1,19**	**1,16**	**1,06**	**1,07**	**1,02**	0,93	0,97	0,91	0,88	0,93	0,91	0,93
Polska	0,17	0,19	0,21	0,23	0,29	0,24	0,35	0,39	0,40	0,43	0,64	0,73	0,79
Słowacja	0,66	0,54	0,55	0,50	0,61	0,53	0,55	0,51	0,54	0,62	0,79	**1,08**	1,00

Źródło: jak przy tabl. 1.

Kolejną grupą były usługi o najmniejszej perspektywie rozwoju. Kategoria ta powstała ze względu na występowanie bardzo niskiej wartości *RCA* dotyczącego konkurencyjności. Do tej grupy zaliczyć można usługi finansowe oraz patenty i licencje.

Analizę tej grupy rozpoczyna przybliżenie sytuacji w eksporcie usług finansowych. Żaden z analizowanych krajów (poza latami 2000 i 2001) nie wykazywał ujawnionej przewagi komparatywnej w międzynarodowym handlu usługami finansowymi. Mimo że Polska wypadła najlepiej wśród krajów Europy Środkowej, to konkurencyjność usług nie była wystarczająca do osiągnięcia specjalizacji. Najbliżej do uzyskania ujawnionej przewagi komparatywnej w obrocie usługami finansowymi były Niemcy, które od 2010 r. charakteryzowały się ciągłym wzrostem wartości *RCA*.

TABL. 10. WARTOŚCI *RCA* W EKSPORCIE ŚWIADCZONYCH USŁUG FINANSOWYCH WEDŁUG KRAJÓW

Kraje	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Czechy	**1,99**	**1,21**	0,55	0,48	0,90	0,66	0,41	0,23	0,12	0,05	0,05	0,06	0,03
Niemcy	**1,75**	**1,58**	0,70	0,69	0,74	0,74	0,68	0,70	0,80	0,83	0,73	0,78	0,81
Polska	0,36	0,77	0,28	0,31	0,25	0,25	0,16	0,18	0,23	0,23	0,25	0,18	0,18
Słowacja	0,58	0,42	0,50	0,37	0,49	0,58	0,40	0,51	0,35	0,77	0,11	0,07	0,10

Źródło: jak przy tabl. 1.

Ostatnią z omawianych kategorii były patenty i licencje. Sytuacja Polski, Słowacji i Czech w tej dziedzinie była bardzo podobna. Niemal przez cały okres badania nasz kraj był na ostatnim miejscu pod względem handlu patentami i licencjami (w niektórych latach kilka razy gorszy wynik uzyskały Czechy). Sytuacja taka wynikała przede wszystkim z bardzo niskiej skali prowadzonych w naszym kraju badań oraz składanych w urzędzie wniosków patentowych, jak również pobierania opłat za używanie znaków towarowych. Najlepiej w tej kategorii prezentowały się Niemcy (w analizowanym okresie wykazywały wartość *RCA* powyżej jedności).

TABL. 11. WARTOŚCI *RCA* W EKSPORCIE PATENTÓW I LICENCJI

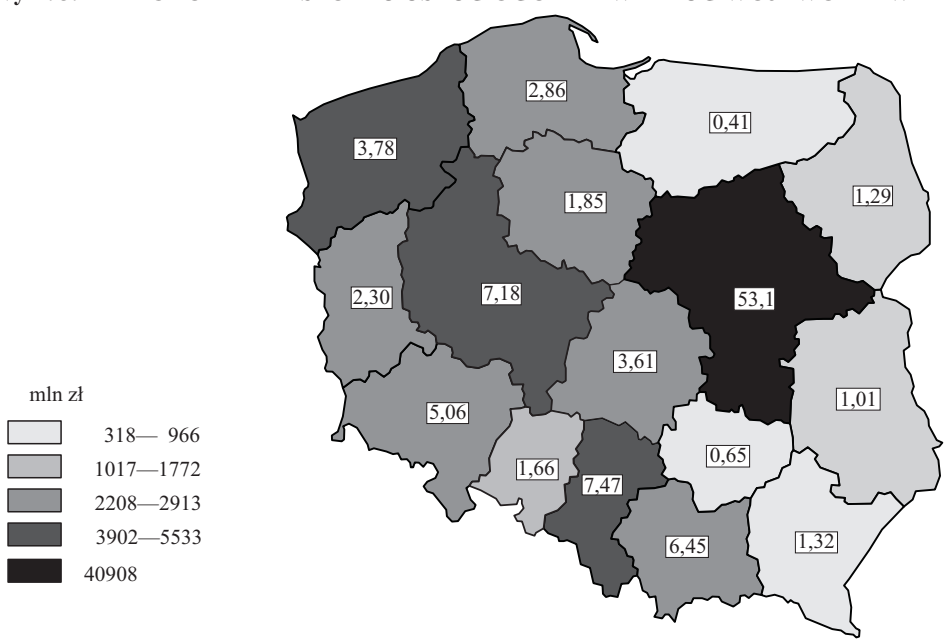
Kraje	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Czechy	0,27	0,24	0,29	0,21	0,12	0,10	0,07	0,06	0,06	0,10	0,09	0,12	0,24
Niemcy	**1,42**	**1,72**	**1,69**	**1,15**	**1,16**	**1,26**	**1,15**	**1,12**	**1,04**	**1,33**	**1,11**	**1,46**	**1,33**
Polska	0,14	0,23	0,15	0,08	0,07	0,11	0,06	0,10	0,16	0,07	0,13	0,20	0,15
Słowacja	0,31	0,32	0,61	0,54	0,50	0,50	0,51	0,63	0,47	0,28	0,14	0,02	0,02

Źródło: jak przy tabl. 1.

PRZYCHODY Z EKSPORTU USŁUG W 2012 R. WEDŁUG WOJEWÓDZTW

Dane wykorzystane w tej części opracowania pochodzą z REGON-u oraz z Badania Międzynarodowego Handlu Usługami (dane bilansu płatniczego). Do analizy wzięto 12531 firm badanych w 2012 r. Badanie nie uwzględnia danych dotyczących usług z tytułu podróży zagranicznych.

Wykr. 3. PRZYCHODY Z EKSPORTU USŁUG OGÓŁEM WEDŁUG WOJEWÓDZTW



Uwaga. Przedziały zaprezentowane w legendzie obrazują przychody z eksportu najwyższe i najniższe w danej grupie województw; na mapie podano udział procentowy w eksporcie usług ogółem.
Źródło: jak przy wykr. 1.

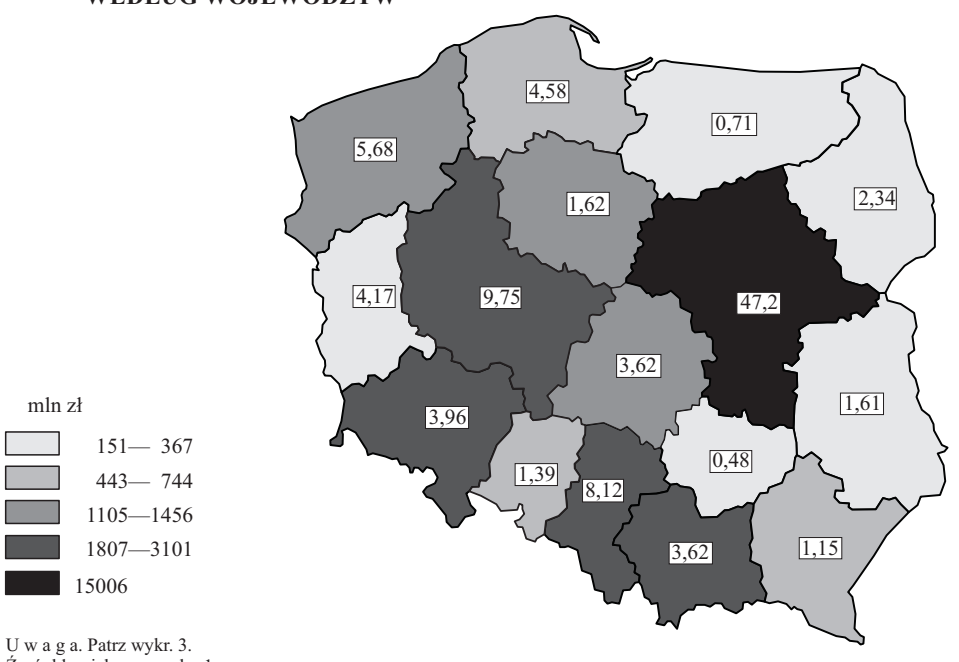
W 2012 r. polski eksport usług osiągnął wartość 77075 mln zł, na którą składały się pozostałe usługi — 45272 mln zł oraz usługi transportowe w kwocie 31803 mln zł. Największe przychody z eksportu usług w prawie wszystkich kategoriach wygenerowały przedsiębiorstwa przypisane do woj. mazowieckiego.

Wynika to z faktu, że wiele firm jako miejsce siedziby deklarowało to województwo, mimo iż faktycznie ich działalność mogła być prowadzona w innej części kraju.

Przychody z eksportu usług firm umiejscowionych w woj. mazowieckim wynosiły 40908 mln zł, co stanowiło 53,1% wszystkich przychodów pochodzących z międzynarodowego handlu usługami. Kolejne miejsca zajmowały następujące województwa: śląskie (7,47%, tj. 5754 mln zł), wielkopolskie (7,18% — 5533 mln zł) i małopolskie (6,45% — 4972 mln zł). Najgorzej pod względem wielkości przychodów radziły sobie firmy na terenach wschodniej oraz północno-wschodniej Polski, ich przychody nie przekraczały 996 mln zł (suma przychodów dla 4 najsłabszych województw[2] wynosiła 2607 mln zł, co stanowiło zaledwie 3,36%). Pokazuje to obecną polaryzację struktury geograficznej w przychodach z międzynarodowego handlu usługami.

Kolejnym etapem analizy było wyodrębnienie głównych kategorii usługowych, tj. usług transportowych oraz pozostałych. Rozmieszczenie w kraju pierwszej z nich obrazuje wykr. 4.

Wykr. 4. PRZYCHODY Z EKSPORTU USŁUG TRANSPORTOWYCH WEDŁUG WOJEWÓDZTW



U w a g a. Patrz wykr. 3.
Ź r ó d ł o: jak przy wykr. 1.

[2] Warmińsko-mazurskie, podlaskie, lubelskie i świętokrzyskie.

W kategorii tej obserwujemy znaczącą dominację w usługach transportowych woj. mazowieckiego (47,2%, tj. 15006 mln zł), kolejne miejsca zajmowały województwa wielkopolskie (9,75% — 3101 mln zł) i śląskie (8,12% — 2581 mln zł).

Bardzo dobry wynik notowało woj. zachodniopomorskie (5,68%, tj. 1806 mln zł). Jak można zauważyć, większe przychody z usług transportowych obserwowano na terenach zachodnich naszego kraju (w stosunku do usług ogółem oraz pozostałych usług). Wynikało to z faktu, że głównymi odbiorcami polskich towarów są kraje UE (przeważają Niemcy).

Największe przychody z eksportu usług transportowych uzyskały firmy z woj. mazowieckiego dzięki dominującej roli w transporcie samochodowym (towarowym), lotniczym oraz kolejowym (w przewozie towarów oraz pasażerów). Wynika to głównie z dogodnego położenia regionu oraz dużej liczby siedzib przedsiębiorstw transportowych. Przychody ze świadczenia omawianych usług na rzecz nierezydentów przez firmy mieszczące się w woj. wielkopolskim składały się głównie z dobrze rozwiniętego transportu towarów (samochodowy, lotniczy), jak też z usług wspomagających transport samochodowy. Kluczową rolę odgrywało również położenie i bliskość rynku niemieckiego.

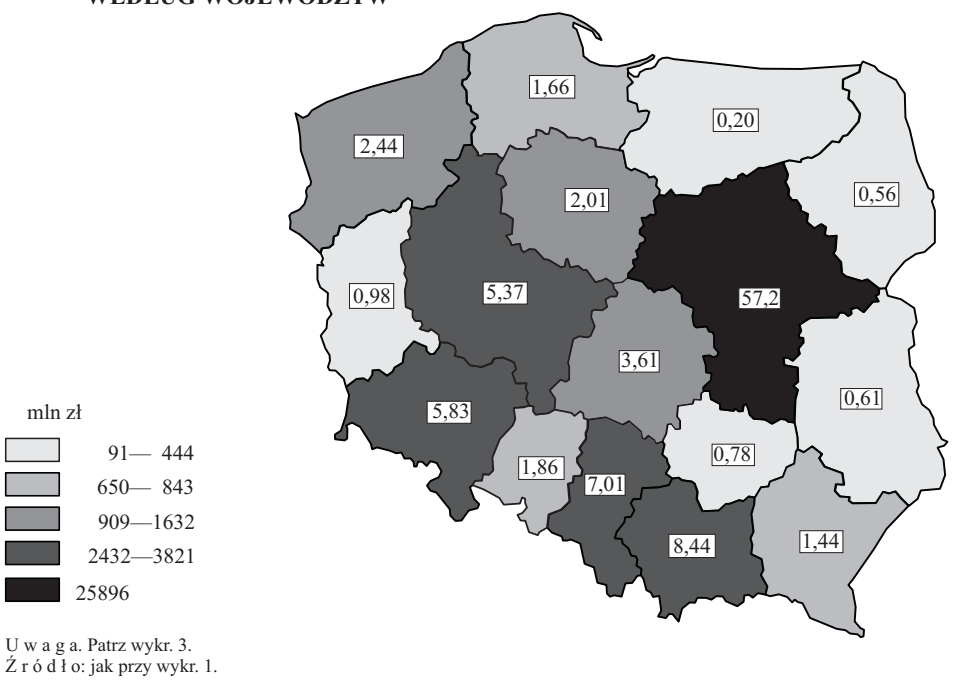
Z kolei na eksport usług transportowych woj. śląskiego miały wpływ głównie usługi przewozu towarów za pomocą samochodów oraz kolei. Nie bez znaczenia dla tego regionu były pozostałe usługi transportu kolejowego i samochodowego. Udział firm z woj. zachodniopomorskiego był związany przede wszystkim z dominującą rolą w transporcie morskim (223 mln zł) oraz żegludze śródlądowej (31 mln zł). Wynik ten nie był zaskoczeniem ze względu na obecny w Szczecinie port wraz ze stocznią, jak również dużą rolę, jaką odegrało sąsiedztwo z Niemcami. W przewozie towarów drogą morską największe przychody notowały przedsiębiorstwa z woj. pomorskiego.

Dobre miejsce w eksporcie usług z tytułu pozostałego transportu samochodowego oraz pozostałych usług transportu lotniczego zajęły firmy z woj. dolnośląskiego: usługi wspomagające transport samochodowy — 16,3% (106 mln zł) oraz usługi wspomagające transport lotniczy — 2,75% (40 mln zł). Warto wspomnieć o sytuacji firm z woj. opolskiego, które w usługach związanych z pasażerskim transportem samochodowym osiągnęły przychody na poziomie 45 mln zł (41,1%). W tym transporcie region ten przewyższył nawet woj. mazowieckie (o 6,1 p.proc.).

Następną omawianą kategorią były pozostałe usługi, których wartość eksportu wyniosła 45270 mln zł. W ich skład wchodzą: usługi pocztowe, budowlane, ubezpieczeniowe, finansowe, patenty i licencje, pozostałe usługi gospodarcze, usługi kulturalne i rekreacyjne oraz usługi rządowe (*Balance…*, 2009). Również w tym przypadku dominującym regionem okazało się woj. mazowieckie, którego przychód z eksportu w tej kategorii stanowił 57,2% (25896 mln zł) całego eksportu pozostałych usług. Kolejne miejsca należały do firm z województw

małopolskiego (8,44%, tj. 3821 mln zł), śląskiego (7,01% — 752 mln zł) i dolnośląskiego (5,83% — 2641 mln zł). Przychody z pozostałych usług pochodziły głównie z centralnego oraz południowego regionu kraju (wykr. 5).

Wykr. 5. PRZYCHODY Z EKSPORTU USŁUG POZOSTAŁYCH WEDŁUG WOJEWÓDZTW



U w a g a. Patrz wykr. 3.
Ź r ó d ł o: jak przy wykr. 1.

Przychody były głównie generowane przez przedsiębiorstwa z siedzibą w województwach centralnych (mazowieckie i wielkopolskie) oraz południowych (dolnośląskie, śląskie i małopolskie). Również w tym przypadku najbardziej wyróżniały się firmy z woj. mazowieckiego (duża liczba firm, dogodne położenie geograficzne, najbogatszy region). Najsłabiej w eksporcie omawianej grupy wypadły województwa położone na wschodzie Polski: podlaskie, lubelskie i warmińsko-mazurskie.

Z grupy pozostałych usług można wydzielić rodzaje usług, których eksport zdominował region mazowiecki. Grupę tę stanowią usługi: ubezpieczeniowe, pocztowe i telekomunikacyjne, finansowe, pozostałe usługi gospodarcze, usługi kulturalne i rekreacyjne oraz usługi informatyczne i informacyjne.

W eksportowanych przez Polskę usługach budowlanych (jedna z polskich dziedzin specjalizacji) bardzo dobrze wypadły firmy pochodzące ze Śląska (16%, tj. 696 mln zł), Wielkopolski (12,3% — 538 mln zł) oraz woj. opolskiego (7,55% — 329 mln zł).

W przychodach z eksportu patentów i licencji dominowały firmy z woj. mazowieckiego, stanowiące 45,3% (295 mln zł) w eksporcie ogółem; kolejnymi regionami były województwa małopolskie (12,5% — 82 mln zł, które wyprzedziło tym samym znane w całym kraju ośrodki badawcze z woj. dolnośląskiego o 1,3 p.proc.) i łódzkie (10,5% — 68 mln zł).

Pierwszą grupę wśród usług finansowych stanowiły firmy z województw: mazowieckiego, małopolskiego, dolnośląskiego i kujawsko-pomorskiego (regiony te łącznie generowały 96,18% całego eksportu omawianej kategorii usług). Świadczy to o znacznej specjalizacji tej grupy województw. Podobna sytuacja miała miejsce w usługach kulturalnych i rekreacyjnych oraz pozostałych usługach gospodarczych.

Grupę pozostałych usług gospodarczych stanowią m.in. usługi: związane z handlem, leasingu operacyjnego, prawne, w zakresie księgowości i rachunkowości, związane z sondażami oraz public relations, architektoniczne i inżynieryjne, badawczo-rozwojowe, utylizacji odpadów, rolnicze i wydobywcze, między przedsiębiorstwami powiązanymi, pozostałe usługi (*Balance…*, 2009). W tej kategorii znów można wyróżnić eksport usług, w których bezkonkurencyjną dominację uzyskało woj. mazowieckie. Były to głównie usługi doradcze świadczone przedsiębiorstwom (prawne, reklama, badania rynku i sondaże opinii publicznej, doradztwo gospodarcze w zakresie public relations, związane z zarządzaniem, usługi badawczo-rozwojowe, związane z handlem oraz leasingiem operacyjnym).

W kontekście usług architektonicznych, inżynieryjnych oraz pozostałych usług technicznych większość przychodów generowana była przez firmy przypisane do województw położonych w centralnej części naszego kraju (mazowieckie i kujawsko-pomorskie) oraz na południu (śląskie i małopolskie).

Pokazuje to, że w grupie najsłabszych można wyróżnić te województwa, które relatywnie lepiej wypadały w międzynarodowej sprzedaży w poszczególnych podgrupach usługowych. Mimo dosyć słabej sytuacji w przychodach uzyskanych z międzynarodowego handlu usługami takich regionów, jak Podlasie i Lubelszczyzna, można jednak wskazać dziedziny dominującej działalności w tym zakresie. Przedsiębiorstwa z woj. podlaskiego posiadały udział rzędu 1,29% w ogólnej sprzedaży usług, przy czym w usługach transportowych było to 2,34%, w pozostałych usługach tylko 0,56%. Podobna sytuacja miała miejsce wśród firm z woj. lubelskiego (udział w przychodach ogółem w międzynarodowym handlu usługami wynosił 1,02%, przy czym udział w usługach transportowych to 1,61%, zaś w przychodach z pozostałych usług 0,61%). Do grupy eksportującej relatywnie więcej usług transportowych dołączyć można firmy z kolejnych dwóch województw. Pierwszym z nich było woj. lubuskie, firmy z tego regionu wykorzystując położenie geograficzne osiągnęły udziały w eksporcie następujących usług: ogółem — 2,3%, transportowe — 4,17%, pozostałe usługi — 0,98%. Drugim wspomnianym regionem, w którym dominowały przedsię-

biorstwa z przeważającą działalnością w transporcie było woj. pomorskie, osiągające następujące udziały w eksporcie usług: 2,86% w ogółem, 4,58% w transportowych oraz 1,66% w pozostałych. Kolejną grupą były regiony świadczące nierezydentom głównie pozostałe usługi. Do ich grona zaliczyć można firmy z siedzibą w woj. kujawsko-pomorskim (udziały w eksporcie stanowiły: 1,85% usług ogółem, 1,62% transportowych oraz 2,01% w pozostałych usługach). Firmy z siedzibą w woj. opolskim charakteryzowały się następującymi udziałami usług: ogółem — 1,66%, transportowe — 1,39% oraz pozostałe — 1,86%. Najsłabiej w tej grupie radziły sobie przedsiębiorstwa z woj. świętokrzyskiego, które osiągnęły następujące udziały usług: ogółem — 0,65%, transportowe — 0,48% i pozostałe — 0,78%.

Wnioski

Zgodnie z założonym celem zidentyfikowano usługi, w których Polska wykazała ujawnioną przewagę komparatywną były to usługi budowlane ($RCA = 1{,}48$), transportowe ($RCA = 1{,}5$) oraz podróże zagraniczne ($RCA = 1{,}34$). Usługi te wymagają nieco mniejszego nakładu wykwalifikowanej siły roboczej.

Wartość bliską jedności osiągnął *RCA* usług kulturalnych i rekreacyjnych ($RCA = 0{,}85$) oraz pozostałych usług gospodarczych ($RCA = 0{,}89$). Obie kategorie usług mają największe perspektywy rozwoju ze względu na obserwowany w *RCA* trend wzrostowy.

Kolejną grupą usług, w której notowano sukcesywną poprawę sytuacji były usługi informatyczne i informacyjne, jednak osiągnięty wynik nie upoważnia do nazwania ich specjalizacją.

Można stwierdzić, że w międzynarodowym handlu usługami nasza gospodarka należy do krajów opartych na usługach pracochłonnych głównie ze względu na specjalizację w usługach transportowych, podróżach zagranicznych oraz usługach budowlanych. Należy zwrócić uwagę na poprawę konkurencyjności w zakresie: usług kulturalnych i rekreacyjnych, informatycznych i informacyjnych oraz w pozostałych usługach gospodarczych należących do usług angażujących kapitał ludzki wysokiej jakości.

Najsłabiej w polskim eksporcie usług wypadł handel usługami związanymi z opłatami z tytułu patentów i licencji oraz usługami finansowymi.

Drugim analizowanym zagadnieniem był eksport usług według województw. Dużą koncentrację tego eksportu notowano w woj. mazowieckim. Jak już wspomniano, wynikało to z umiejscowienia głównych siedzib firm prowadzących międzynarodowy handel usługami w Warszawie. Ponadto woj. mazowieckie należy do najbogatszych regionów naszego kraju, jak również nie bez znaczenia jest dogodne położenie geograficzne i większa obecność wyspecjalizowanego personelu.

Kolejne miejsca zajmowały z reguły województwa: śląskie, wielkopolskie, małopolskie oraz dolnośląskie — regiony o stosunkowo dużej powierzchni, które można również określić jako regiony bogate i ze znaczącą liczbą specjalistów. Najniższe przychody generowane były przez przedsiębiorstwa z siedzibą w województwach: lubelskim, podlaskim, warmińsko-mazurskim oraz świętokrzyskim.

Stwierdzono wystepowanie dwóch grup województw specjalizujących się w eksporcie usług. Pierwsza z nich skupia województwa, w których usługi transportowe przeważają w eksporcie usług ogółem. Do tej grupy zaliczyć można przede wszystkim województwa: wielkopolskie, śląskie, zachodnio-pomorskie oraz lubuskie; ich rezultaty osiągnięte w eksporcie usług transportowych wynikają głównie z dobrego położenia geograficznego. Drugą grupę stanowiły firmy dostarczające pozostałe usługi dla nierezydentów z przewagą tej kategorii w eksporcie usług ogółem. Do tej grupy zaliczyć można przede wszystkim woj. małopolskie i dolnośląskie, położone w regionie bogatym, z dużą liczbą wykwalifikowanej kadry pracowniczej. Również w tym przypadku dominującym regionem okazało się woj. mazowieckie, kolejne miejsca należały do województw małopolskiego, śląskiego i dolnośląskiego. Przychody z pozostałych usług pochodziły głównie z centrum oraz południa kraju.

mgr Łukasz Matuszczak — *NBP*

LITERATURA

Balance of Payments and International Investment Position Manual, Sixth Edition (2009), International Monetary Fund, IMF Multimedia Services Division

Balassa B. (1986), *Comparative advantage in manufactured goods: a reappraisal*, The Review of Economics and Statistics

Hagemejer J., Michałek J. J., Michałek T. (2006), *Pogłębienie rynku wewnętrznego UE: skutki liberalizacji w sektorze usług. Potencjalne znaczenie budżetu UE jako narzędzia łagodzenia zmian strukturalnych związanych z liberalizacją sektora usług*

Manual of Statistics of International Trade in Service 2010 (2011), Department of Economic and Social Affairs, United Nations

Mongiało D. (2007), *Specjalizacja eksportowa krajów UE w międzynarodowym handlu usługami*, Ministerstwo Gospodarki

SUMMARY

Two related problem are considered. The first concerns a situation in Polish international trade in services. It shows conclusions obtained from analyzing, proposed by B. Balassa, Revealed Comparative Advantage (RCA) of the exports of services. Data time series included information from period 2000—2012.

Comparison between Poland and neighbors economy (Germany, Czech and Slovakia) was done. The second concerns the geographical breakdown of the exports of services data. This breakdown was made from whole economy into voivodship. Specialization of each group of voivodships was analyzed.

РЕЗЮМЕ

В статье был представлен анализ обследования касающегося формирования выявленного сравнительного преимущества и географической структуры предприятий предоставляющих услуги за границу. Была охарактеризована динамика и структура услуг по группам, а также была обследована конкурентоспособность польских услуг. В анализе была использована мера RCA (revealed comparative advantage) предложенная Б. Балласом. Уровень конкурентоспособности экспорта услуг в Польше был сопоставлен с соседними странами (Германией, Чешской Республикой и Словакей). Было проанализировано расположение по воеводствам предприятий предоставляющих услуги на экспорт. Статья характризует также специализацию воеводств в области экспорта конкретных подгрупп услуг.

INFORMACJE. PRZEGLĄDY. RECENZJE

Historia Polski w liczbach — unikalna seria wydawnicza statystyczno-historyczna GUS

Dzieje Polski w badaniach historycznych po II wojnie światowej przedstawiano głównie w kontekście politycznym oraz społeczno-gospodarczym. Brakowało natomiast wydawnictw pozwalających odtworzyć przeszłość, od najstarszych dziejów naszego kraju aż po dzień dzisiejszy, w ujęciu statystycznym.

Inicjatywę takiego opracowania podjął GUS w końcu lat 60. XX w. Efektem tych prac było przygotowanie w wersji roboczej kilkunastu zeszytów pt. *Historyczny Rocznik Statystyczny* — o małej objętości i skromnym zestawie tematycznym — zawierających dane dla lat 1918—1968. Opracowanie to (powielone w niewielkiej liczbie egzemplarzy) przekazano jedynie wąskiemu gronu osób w celu zebrania opinii dotyczących przedstawiania problematyki badań historycznych w liczbach.

Podjęte kolejne prace nad wydaniem *Historycznego Rocznika Statystycznego*, który w założeniu miał być udostępniony powszechnie, jednak nie zostały dokończone, głównie z powodów finansowych. Materiał statystyczny zgromadzony wówczas w dużej mierze wykorzystano później w wydawnictwach GUS: *Polska 1918—1978*, wydanie z 1978 r. i *Polska 1918—1988*, wydanie z 1989 r. Druga publikacja ukazała się z inicjatywy ówczesnego prezesa GUS dra Franciszka Kubiczka.

Jednocześnie w II połowie 1989 r. wznowiono prace nad publikacją dokumentującą w pełnym zakresie rozwój społeczno-gospodarczy ziem polskich. W celu jej opracowania prezes Franciszek Kubiczek powołał 31 lipca 1989 r. Zespół Redakcyjny publikacji *Historia Polski w liczbach*, któremu przewodniczy do chwili obecnej. Jednocześnie rozszerzono zakres chronologiczny i tematyczny publikacji na okres przedrozbiorowy i czas rozbiorów.

Redaktorem głównym był prof. dr hab. Andrzej Jezierski z Wydziału Nauk Ekonomicznych Uniwersytetu Warszawskiego. Opracował on wraz z Franciszkiem Kubiczkiem koncepcję całości wydawnictwa. Po śmierci prof. Jezierskiego w 2002 r. Zespołem Redakcyjnym kierował prof. dr hab. Andrzej Wyczański z PAN, po nim prof. dr hab. Juliusz Łukasiewicz z Instytutu Historycznego Uniwersytetu Warszawskiego. Konsultantem merytorycznym była dr Halina Dmochowska — wiceprezes GUS.

W skład Zespołu Redakcyjnego weszli — poza osobami wspomnianymi — prof. dr hab. Cezary Kuklo z Instytutu Historii Uniwersytetu w Białymstoku, dr Cecylia Leszczyńska z Wydziału Nauk Ekonomicznych Uniwersytetu War-

szawskiego i dr Andrzej Radźko z wydawnictwa „Pallatinum". Funkcję sekretarza pełnił mgr Jan Berger, a od 2009 r. również mgr Antoni Żurawicz — pracownicy GUS. Prace redakcyjne wykonywała mgr Grażyna Szydłowska z GUS.

Zespół ten w pierwszym etapie swych prac w latach 1990—1999 opracował 9 zeszytów tematycznych. Były to: *Ludność, Terytorium*; *Rolnictwo, Leśnictwo*; *Górnictwo i Przemysł, Budownictwo, Dochód Narodowy*; *Oświata, Nauka, Kultura*; *Transport, Łączność*; *Handel*; *Finanse; Materialne warunki życia ludności*; *Państwo, Wojsko*. Ponadto w 1994 r. wydano publikację pt. *Economic History of Poland in Numbers* zawierającą wybór najważniejszych danych statystyczno-historycznych w języku angielskim.

Zebrane w tych zeszytach materiały statystyczne posłużyły do opracowania nowej wersji publikacji *Historia Polski w liczbach* w dwóch tomach: tom I — *Państwo, Społeczeństwo* wydany w 2003 r. i tom II — *Gospodarka* opublikowany w 2006 r. W obu tomach, jak i w późniejszych publikacjach tej serii wyodrębniano trzy okresy chronologiczne: do 1795 r., lata 1795—1918 i okres 1918—2000 (w dalszych opracowaniach uwzględniano kolejne lata XXI w.). Terytorium państwa polskiego w obu tomach pokazano w ujęciu statystycznym dla takiego obszaru państwa, jaki obejmowały jego granice w danym okresie. Dla lat 1795—1918 starano się podawać dane statystyczne według tych obszarów, które uważano w okresie I wojny światowej za terytorium przyszłego państwa polskiego na podstawie granic historycznych, na których zamieszkiwały zwarte grupy Polaków.

W 2009 r. Zespół Redakcyjny został reaktywowany przez prezesa GUS prof. dra Józefa Oleńskiego, który włączył do jego składu mgr Bożenę Łazowską — dyrektora Centralnej Biblioteki Statystycznej i mgra Władysława Wiesława Łagodzińskiego — rzecznika prezesa GUS. Zespół opracował popularno-naukowy tom pt. *Zarys historii Polski w liczbach* opublikowany w 2012 r. Zawierał on bogaty wybór tablic ze wspomnianego wyżej dwutomowego opracowania, uzupełniony nowymi materiałami statystycznymi. Tablice prezentowały dane dotyczące m.in.: terytorium, władzy centralnej i terenowej, ludności, narodowości i wyznań, aktywności ekonomicznej ludności, cen, płac i spożycia, ochrony zdrowia, edukacji i kultury, a także rolnictwa i rybołówstwa, przemysłu i budownictwa, handlu i dochodu narodowego.

W 2014 r., w 10. rocznicę wstąpienia Polski do Unii Europejskiej (UE), Zespół Redakcyjny przygotował publikację pt. *Polska w Europie*, stanowiącą tom III serii „Historia Polski w liczbach". Autorami merytorycznymi, którzy podjęli się opracowania tablic dotyczących określonych dziedzin życia społecznego i gospodarczego w poszczególnych okresach historycznych Polski i krajów europejskich byli: prof. dr hab. Cezary Kuklo — okres do 1795 r., prof. dr hab. Juliusz Łukasiewicz — lata 1795—1918 i dr Cecylia Leszczyńska — lata 1918—2010. W swych opracowaniach przedstawili przebieg naj-

ważniejszych procesów i wydarzeń społeczno-gospodarczych zachodzących na przestrzeni wieków w Polsce na tle innych krajów europejskich. Każde z tych opracowań poprzedzono szczegółowymi uwagami metodologicznymi i komentarzami historycznymi.

Publikacja składa się z rozdziałów tematycznych dotyczących: terytorium i ludności, narodowości i wyznań, materialnych warunków życia ludności, edukacji i kultury oraz sportu, rolnictwa, przemysłu, transportu i łączności, handlu zagranicznego, finansów, dochodu narodowego, a także oddzielnych tablic pokazujących współczesną Polskę wśród krajów UE. Większość tablic zbudowano na podstawie źródeł pierwotnych, jakimi są roczniki statystyczne oraz publikacje instytucji odpowiedzialnych za gromadzenie i przetwarzanie danych. Podstawą opracowania stały się publikacje GUS, Ligi Narodów, ONZ, Eurostatu, Banku Światowego oraz opracowań międzynarodowych organizacji statystycznych i gospodarczych.

Autorzy tego tomu w znacznym zakresie korzystali także z dzieł B. R. Michella pt. *International Historial Statistics* (5 tomów, 6 wydań — ostatnie w 2007 r.) oraz Agnusa Maddisona dotyczących światowej gospodarki, wydanych przez OECD w latach 2003 i 2007.

Publikację przygotowano w wersji polsko-angielskiej. Opracowanie edytorskie z piękną szatą graficzną zapewnił Zakład Wydawnictw Statystycznych GUS.

Program dalszych prac serii „Historia Polski w liczbach" zakończy tom IV, który poświęcony będzie problematyce historii instytucji statystycznych w Polsce i prowadzonych przez nie badań.

oprac. **Jan Berger**

Ogólnopolska konferencja użytkowników oprogramowania statystycznego R

W dniach 15—17 października 2014 r. na Uniwersytecie Ekonomicznym w Poznaniu odbyła się pierwsza ogólnopolska konferencja pt. *Polski Akademicki Zlot Użytkowników R*. Organizatorami konferencji była Katedra Statystyki Uniwersytetu Ekonomicznego w Poznaniu, Katedra Metod Matematycznych i Statystycznych Uniwersytetu Przyrodniczego w Poznaniu oraz Studenckie Koło Naukowe „Estymator" działające przy Katedrze Statystyki UEP. Honorowy patronat nad konferencją objęli prof. dr hab. Emil Panek, dziekan Wydziału Informatyki i Gospodarki Elektronicznej UEP, prof. dr hab. Wiesław Koziara, dziekan Wydziału Rolnictwa i Bioinżynierii UP oraz dr Jacek Kowalewski, dyrektor Urzędu Statystycznego w Poznaniu. Patronat nad konferencją objął również oddział poznański PTS i Polskie Towarzystwo Biometryczne. Konferencja sponsorowana była przez firmy Revolution Analytics i Analyx oraz fundację SmarterPoland.pl.

Celem konferencji było przedstawienie możliwości wykorzystania pakietu statystycznego R w różnych dziedzinach nauki oraz biznesu, wymiana doświadczeń oraz integracja środowiska użytkowników pakietu statystycznego R. Pierwszy dzień konferencji poświęcony był warsztatom z zakresu analizy i eksploracji danych przestrzennych, przetwarzania dużych zbiorów danych z pakietem data.table, analizy danych sondażowych z pakietem survey, wizualizacji oraz wprowadzenia do analizy danych w pakiecie statystycznym R. W warsztatach wzięło udział blisko 130 osób.

Drugiego i trzeciego dnia konferencji przedstawiono 29 referatów poświęconych różnym aspektom zastosowania pakietu statystycznego R w naukach przyrodniczych, ekonomicznych oraz informatycznych, jak również aplikacjom biznesowym. Udział w konferencji był bezpłatny, co sprzyjało wyjątkowej frekwencji. W warsztatach i konferencji uczestniczyło ponad 100 osób, nauczycieli akademickich reprezentujących różne dyscypliny naukowe oraz przedstawicieli firm i instytucji zainteresowanych praktycznym wykorzystaniem pakietu R. W ramach konferencji wygłoszono dwa zamówione referaty — dr hab. Katarzyny Kopczewskiej (Wydział Ekonomii UW) pt. *Przestrzenne modelowanie zjawisk ekonomicznych w R* oraz dra hab. Przemysława Biecka (Interdyscyplinarne Centrum Modelowania Matematycznego i Informatycznego UW) pt. *Projekt Beta i Bit, diagramy ULAM, wizualizacja z R, czyli o moim analitycznym atelier*.

Oprócz tych dwóch wystąpień uczestnicy konferencji wygłosili 27 referatów (kolejność alfabetyczna): Adolfo Alvarez (Analyx) — *Mastering Data Frames with dplyr*; Zbigniew Bartkowiak (Muzeum Archeologiczne w Poznaniu) — *Wpływ seigerowania i obrzynania półgroszków Władysława Jagiełły na rozkład ich wagi*; Maciej Beręsewicz (Uniwersytet Ekonomiczny w Poznaniu) — *Staty-*

styka małych obszarów w R; Paweł Budzianowski (Uniwersytet im. Adama Mickiewicza) — *Analiza zależności wielowymiarowych za pomocą kopuli w R*; Damian Chmura (Akademia Humanistyczno-Techniczna w Bielsku-Białej) — *Wykorzystanie pakietu R w nauczaniu biocenologii na kierunkach ochrona środowiska i inżynieria środowiska*; Andrzej Dudek (Uniwersytet Ekonomiczny we Wrocławiu) — *Nie tylko CSV i wiersz poleceń — integracja R i aplikacji biznesowych*; Marcin K. Dyderski, Anna K. Gdula (Sekcja Botaniczna Koła Leśników Uniwersytetu Przyrodniczego w Poznaniu) — *Analiza roślinności za pomocą pakietów vegan i Eh of*; Marek Gągolewski (Instytut Badań Systemowych PAN) — *Przetwarzanie napisów przy użyciu pakietu stringi i bibliotek ICU*; Jarosław Jasiewicz (Uniwersytet im. Adama Mickiewicza) — *GIS-R — integracja R z Systemami Informacji Geograficznej*; Paweł Kliber (Uniwersytet Ekonomiczny w Poznaniu) — *R w analizie finansowej*; Paweł Kleka (Uniwersytet im. Adama Mickiewicza) — *Od danych do publikacji. Automatyczne formatowanie zapisu wyników korzystając z pakietu knitr*; Krzysztof Krawiec (Programowanie genetyczne) — *Prog-R-amowanie genetycze*; Marcin Kosiński (Politechnika Warszawska) — *Archiwizowanie artefaktów z pakietem Archivist*; Jakub Nowosad (Uniwersytet im. Adama Mickiewicza) — *Wizualizacja danych przestrzennych w R — od mapy statycznej do interaktywnej*; Michał Okoniewski (ETH Zurich, Scientific IT Services) — *R z zewnętrznymi źródłami danych w genomice — Nonsense case study*; Adrian Olszewski (KCR S.A.) — *GNU R w badaniach klinicznych — czyli warsztat pracy biostatystyka*; Natalia Reszka (PBS sp. z.o.o.) — *Integracja R i NetLogo w symulacjach systemów agendowych*; Kamil Rotter (SKN Estymator, Uniwersytet Ekonomiczny w Poznaniu) — *Internet jako źródło danych na przykładzie serwisu Allegro*; Artur Schwałko (QuantUp) — *Czy/jak R może zastąpić Excela w firmie?*; Agnieszka Strzałka (Uniwersytet Wrocławski) — *Wykorzystanie ggplot2 do wizualizacji wzrostu Streptomyces coelicolor*; Alicja Szabelska, Joanna Zyprych-Walczak (Uniwersytet Przyrodniczy w Poznaniu) — *Wykorzystanie R w analizie danych RNA-seq*; Monika Szczerba, Marek Wiewiórka (Politechnika Warszawska) — *R + big (bio)data warehouses*; Łukasz Wawrowski (Uniwersytet Ekonomiczny w Poznaniu) — *Analiza danych z serwisów społecznościowych*; Marek Wiewiórka (Politechnika Warszawska) — *Scala(-ble) R — integracja środowiska R i języka Scala*; Marek Wiewiórka (Politechnika Warszawska), Alicja Szabelska (Uniwersytet Przyrodniczy w Poznaniu), Michał Okoniewski (ETH Zurich, Scientific IT Services) — *Sygnatury genowe chorób — użycie równoległego R i uczenia maszynowego z MLInterfaces*; Adam Zadgański (Politechnika Wrocławska/QuantUp) — *Analiza i prognozowanie szeregów czasowych w R*; Zygmunt Zawadzki, Daniel Kosiorowski (Uniwersytet Ekonomiczny w Krakowie) — *Odporna eksploracja danych z wykorzystaniem pakietu depthproc*.

Wszystkie prezentacje przedstawione podczas konferencji są dostępne na stronie internetowej www.estymator.ue.poznan.pl/PAZUR.

oprac. **Maciej Beręsewicz**

Wydawnictwa GUS — grudzień 2014 r.

Z grudniowej oferty wydawniczej GUS wśród opracowań cyklicznych warto zwrócić uwagę na następujące pozycje: „Trzeci sektor w Polsce. Stowarzyszenia, fundacje, społeczne podmioty wyznaniowe, samorząd zawodowy i gospodarczy oraz organizacje pracodawców w 2012 r." oraz „Żegluga śródlądowa w Polsce w latach 2010—2013".

Pierwsza z nich **„Trzeci sektor w Polsce. Stowarzyszenia, fundacje, społeczne podmioty wyznaniowe, samorząd zawodowy i gospodarczy oraz organizacje pracodawców w 2012 r."** — wydawana z częstotliwością dwuletnią — prezentuje wyniki badań potencjału społeczno-ekonomicznego trzeciego sektora oraz jego rolę w usługach społecznych i w budowaniu kapitału społecznego. Publikacja powstała jako odpowiedź na wzrastające zainteresowanie problematyką społeczeństwa obywatelskiego. Jest to już drugie opracowanie kompleksowo opisujące wyniki badania „Fundacje i stowarzyszenia oraz jednostki organizacyjne Kościoła katolickiego, innych kościołów i związków wyznaniowych" oraz „Organizacje pracodawców oraz samorządu gospodarczego i zawodowego". Jednym z głównych celów publikacji jest dostarczenie informacji pozwalających na ocenę realizacji polityki publicznej dotyczącej wspierania gospodarki społecznej oraz kapitału społecznego.

Publikacja o charakterze analitycznym poprzedzona jest rozdziałem metodologicznym, w którym zaprezentowano badaną zbiorowość, zastosowane narzędzia, sposób realizacji badań oraz metody uogólniania wyników. W sześciu rozdziałach analitycznych scharakteryzowano sytuację dominującej części organizacji pozarządowych objętych ustawą o działalności pożytku publicznego i o wolontariacie. Czytelnicy znajdą w nich informacje dotyczące m.in. liczebności i struktury aktywnych organizacji, charakterystykę ekonomicznego i społecznego wymiaru działalności, specyfikę jednostek posiadających status organizacji pożytku publicznego, a także przegląd kluczowych zmian, jakie zaszły w trzecim sektorze w latach 1997—2012. Część analityczna została wzbogacona grafiką w postaci tablic, wykresów oraz map.

Publikacja wydana została w formie tradycyjnej w polskiej wersji językowej, dostępna jest również na stronie internetowej Urzędu, a także na płycie CD. Wychodząc naprzeciw oczekiwaniom Czytelników, do publikacji dołączono szeroki zestaw tablic wynikowych w formacie MS Excel, które mogą być wykorzystywane do dalszych analiz.

Publikacja **„Żegluga śródlądowa w Polsce w latach 2010—2013”,** ukazująca się co 4 lata, jest kontynuacją poprzedniej edycji zrealizowanej przez zespół Ośrodka Statystyki Transportu i Łączności Urzędu Statystycznego w Szczecinie. Opracowanie zawiera informacje dot. żeglugi śródlądowej, istotne dla tworzenia i realizacji wspólnej polityki transportowej prowadzonej przez Unię Europejską (UE).

Publikacja składa się z części metodologicznej, komentarza analitycznego oraz tablic wynikowych prezentujących wyniki badań polskiej statystyki publicznej i Urzędu Statystycznego Wspólnot Europejskich (Eurostat). Część metodologiczna obejmuje definicje pojęć stosowanych w statystyce dotyczącej żeglugi śródlądowej. Podstawowym założeniem leżącym u podstaw tej publikacji jest dostarczenie użytkownikom danych statystycznych o drogach wodnych śródlądowych, taborze, krajowych i międzynarodowych przewozach ładunków i pasażerów oraz relacjach ekonomicznych i nakładach inwestycyjnych w polskich przedsiębiorstwach żeglugi śródlądowej. W publikacji uwzględniono również informacje o wynikach finansowych, inwestycjach, zatrudnieniu i wynagradzaniu przez podmioty świadczące usługi w zakresie żeglugi śródlądowej. Dodatkowo publikację wzbogacono o dane z zakresu żeglugi śródlądowej w krajach UE oraz ekonomicznych aspektów ochrony środowiska w odniesieniu do gospodarki wodnej.

Publikacja wydana została w formie tradycyjnej w wersji polsko-angielskiej, dostępna jest również na stronie internetowej Urzędu, a także na płycie CD.

W grudniu 2014 r. ukazały się również: **„Rocznik Statystyczny Rzeczypospolitej Polskiej 2014”, „Bilansowe wyniki finansowe podmiotów gospodarczych w 2013 r.”, „Biuletyn Statystyczny nr 11/2014”, „Budownictwo mieszkaniowe. I—III kwartał 2014 r.”, „Ceny w gospodarce narodowej. Listopad 2014 r.”, „Ceny robót budowlano-montażowych i obiektów budowlanych — październik 2014 r.”, „Działalność gospodarcza podmiotów z kapitałem zagranicznym w 2013 r.”, „Handel zagraniczny. I—IX 2014 r.”, „Informacja o sytuacji społeczno-gospodarczej kraju w listopadzie 2014 r.”, „Informacja o sytuacji społeczno-gospodarczej województw Nr 3/2014”, „Kultura w 2013 r.”, „Leśnictwo 2014”, „Monitoring banków 2013”, „Nakłady i wyniki przemysłu w I—III kwartale 2014 r.”, „Ochrona środowiska 2014”, „Polski rynek ubezpieczeniowy 2013”, „Produkcja i handel zagraniczny produktami rolnymi w 2013 r.”, „Produkcja ważniejszych wyrobów przemysłowych XI 2014 r.”, „Środki trwałe w gospodarce narodowej w 2013 r.”, „Wiadomości Statystyczne Nr 11 — listopad 2014 r.”, „Wiadomości Statystyczne Nr 12 — grudzień 2014 r.”** oraz **„Zużycie paliw i nośników energii w 2013 r.”.**

Oprac. **Justyna Gustyn**

Informacja o sytuacji społeczno-gospodarczej kraju w listopadzie 2014 r.

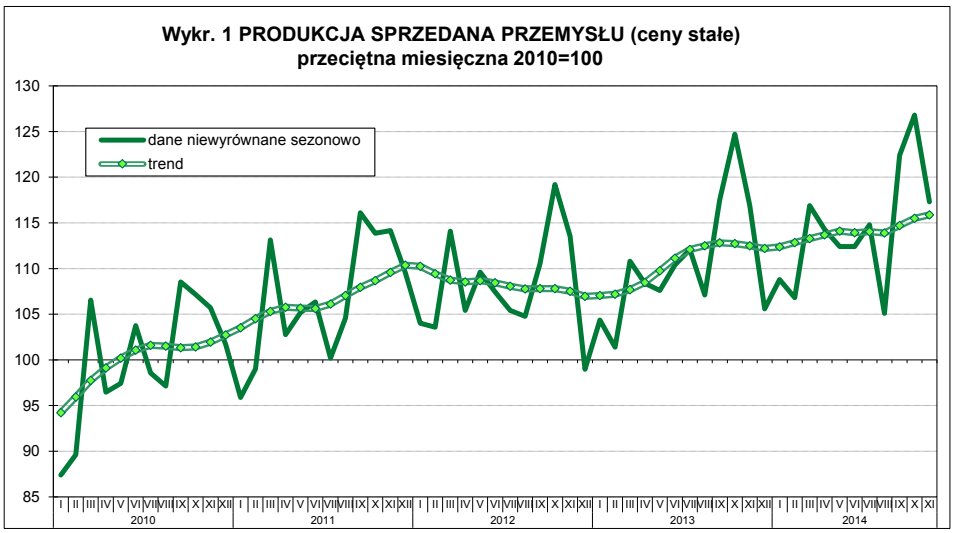
Od IV kwartału 2013 r. obserwuje się umiarkowane, dość stabilne tempo wzrostu gospodarczego. Dynamika w poszczególnych obszarach gospodarki w kolejnych okresach była jednak zróżnicowana. W okresie październik—listopad 2014 r. produkcja sprzedana przemysłu ukształtowała się na poziomie tylko nieco wyższym niż w 2013 r., a produkcja budowlano-montażowa — obniżyła się. Sprzedaż detaliczna w okresie październik—listopad 2014 r. wzrosła w podobnym tempie, jak w trzecim kwartale 2014 r., ale wolniejszym niż w I półroczu; słabsza niż w kolejnych trzech kwartałach 2014 r. była również dynamika sprzedaży usług w transporcie.

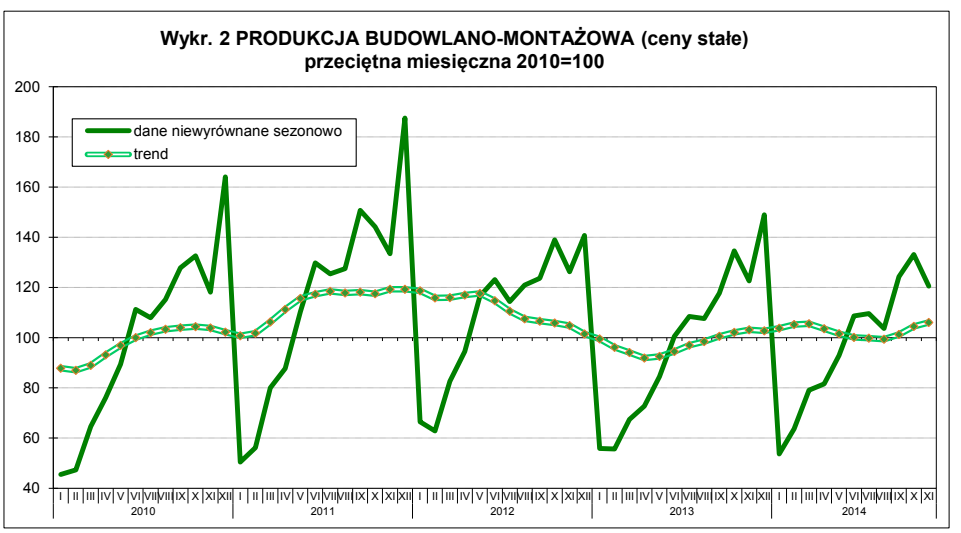
Ceny towarów i usług konsumpcyjnych w okresie styczeń—listopad 2014 r. były tylko nieco wyższe niż w 2013 r. W kolejnych okresach ich dynamika stopniowo słabła i od lipca 2014 r. ceny konsumpcyjne kształtowały się na poziomie niższym niż rok wcześniej. Wpływał na to m.in. spadek cen żywności oraz cen w zakresie transportu. W okresie jedenastu miesięcy 2014 r. utrzymywał się zapoczątkowany w końcu 2012 r. spadek cen producentów w przemyśle i budownictwie w skali roku.

Przy wyższym niż przed rokiem wzroście przeciętnych miesięcznych wynagrodzeń brutto w sektorze przedsiębiorstw i niskiej inflacji, w okresie jedenastu miesięcy 2014 r. obserwowano wzrost w skali roku siły nabywczej płac. Nominalne i realne świadczenia emerytalno-rentowe w obu systemach były wyższe niż w okresie styczeń—listopad 2013 r., choć rosły wolniej niż wynagrodzenia. Lepiej niż w 2013 r. kształtowała się sytuacja na rynku pracy — zatrudnienie nieco wzrosło (wobec spadku w 2013 r.), a bezrobocie rejestrowane obniżyło się.

W listopadzie 2014 r. tempo wzrostu produkcji sprzedanej przemysłu w skali roku było wolniejsze niż w dwóch poprzednich miesiącach i wyniosło 0,3% (po wyeliminowaniu czynników o charakterze sezonowym — 0,2%) (wykr. 1). Zwiększyła się produkcja w przetwórstwie przemysłowym oraz dostawie wody; gospodarowaniu ściekami i odpadami; rekultywacji, przy spadku w pozostałych sekcjach. Spośród głównych grupowań przemysłowych w największym stopniu wzrosła produkcja dóbr konsumpcyjnych trwałych. W okresie styczeń—listopad 2014 r. produkcja sprzedana przemysłu wzrosła w porównaniu z analogicznym okresem 2013 r. o 3,0%. Produkcja budowlano-montażowa w listopadzie 2014 r., w drugim miesiącu z kolei, obniżyła się w skali roku (o 1,6%, a po wyeliminowaniu czynników sezonowych — o 1,1%) (wykr. 2). W okresie styczeń—listopad 2014 r. notowano wzrost o 3,1%, po głębokim spadku obserwowanym w analogicznym okresie 2013 r. Tempo wzrostu sprzedaży detalicznej w listo-

padzie było wolniejsze niż w poprzednich miesiącach i wyniosło 1,4%. W okresie styczeń—listopad 2014 r. sprzedaż zwiększyła się w skali roku o 4,2%.

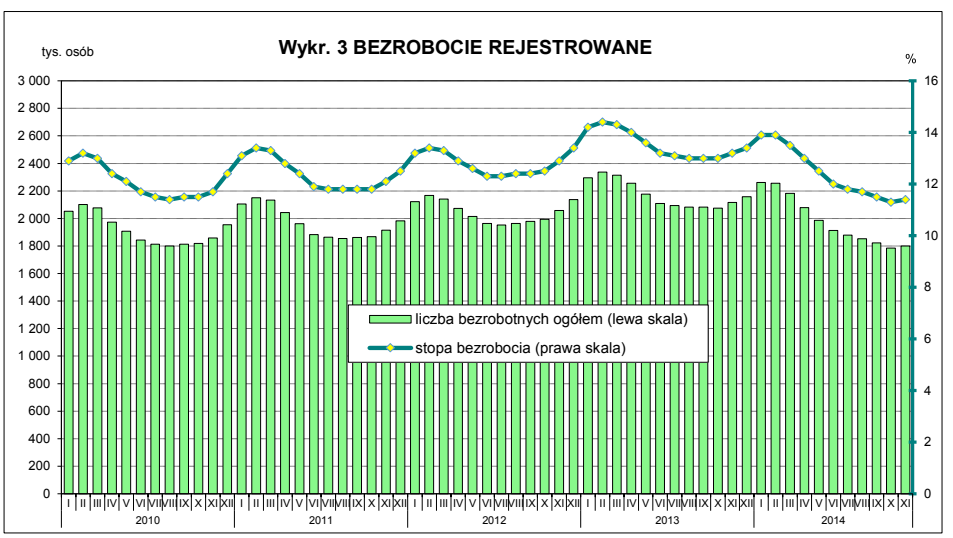




Według badań koniunktury gospodarczej przeprowadzonych w grudniu 2014 r. wskazania dotyczące ogólnego klimatu koniunktury w przetwórstwie przemysłowym, budownictwie oraz handlu detalicznym są gorsze niż przed miesiącem, na co wpływają m.in. czynniki o charakterze sezonowym. Przedsiębiorstwa działające w przetwórstwie przemysłowym oceniają koniunkturę nieznacznie nega-

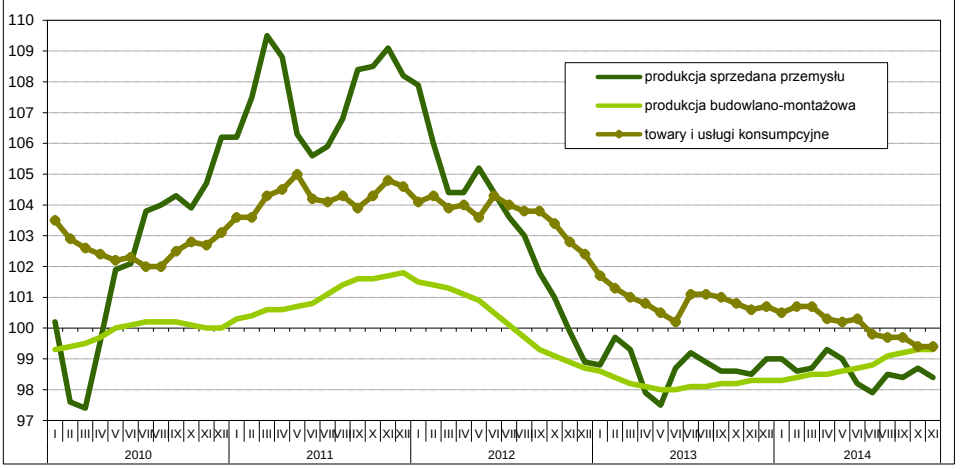
tywnie (wobec pozytywnych ocen w poprzednich jedenastu miesiącach). Niekorzystne, po pozytywnych w listopadzie, są oceny bieżącej sytuacji finansowej, portfela zamówień oraz produkcji, a prognozy w tych obszarach są bardziej negatywne. Pogorszyły się pesymistyczne oceny koniunktury formułowane przez jednostki prowadzące działalność budowlaną. Bardziej niekorzystnie niż przed miesiącem postrzegają one bieżącą i przyszłą sytuację finansową, portfel zamówień oraz produkcję. W handlu detalicznym ogólny klimat koniunktury oceniany jest nieznacznie pesymistycznie, gorzej niż w listopadzie 2014 r. Przewidywania w zakresie popytu na towary oraz diagnozy i prognozy sprzedaży i sytuacji finansowej są niekorzystne, gorsze od zgłaszanych przed miesiącem. Przedsiębiorstwa z tych sekcji zapowiadają dalsze redukcje zatrudnienia (najbardziej znaczące w budownictwie), jak również spadek cen (który w budownictwie i przetwórstwie przemysłowym może być nieco większy niż przed miesiącem). W grudniu 2014 r. poprawiły się natomiast nastroje konsumenckie — zarówno bieżący jak i wyprzedzający wskaźnik ufności konsumenckiej są mniej pesymistyczne niż przed miesiącem.

Na rynku pracy w listopadzie 2014 r. utrzymywały się pozytywne tendencje. Wzrost przeciętnego miesięcznego zatrudnienia w sektorze przedsiębiorstw w skali roku był nieco szybszy niż w poprzednich miesiącach. W okresie jedenastu miesięcy 2014 r. przeciętne zatrudnienie w sektorze przedsiębiorstw wzrosło w skali roku o 0,6%. Stopa bezrobocia rejestrowanego w końcu listopada 2014 r. wyniosła 11,4% (o 1,8 p.proc. mniej niż rok wcześniej) (wykr. 3). Według wyników badania popytu na pracę, w okresie trzech kwartałów 2014 r. utworzono więcej miejsc pracy niż w analogicznym okresie 2013 r., na co wpłynął głównie wzrost odnotowany w III kwartale 2014 r. Znacznie zmniejszyła się liczba zlikwidowanych miejsc pracy.



W listopadzie 2014 r. spadek cen towarów i usług konsumpcyjnych w skali roku był zbliżony do obserwowanego przed miesiącem. W większym stopniu niż w październiku obniżyły się ceny towarów i usług związanych z transportem oraz żywności i napojów bezalkoholowych. Po spadku w poprzednim miesiącu, wzrosły ceny w zakresie łączności. W okresie styczeń—listopad 2014 r. ceny towarów i usług konsumpcyjnych były o 0,1% wyższe niż przed rokiem (wykr. 4). W kolejnych miesiącach utrzymywał się spadek cen producentów w przemyśle i budownictwie.

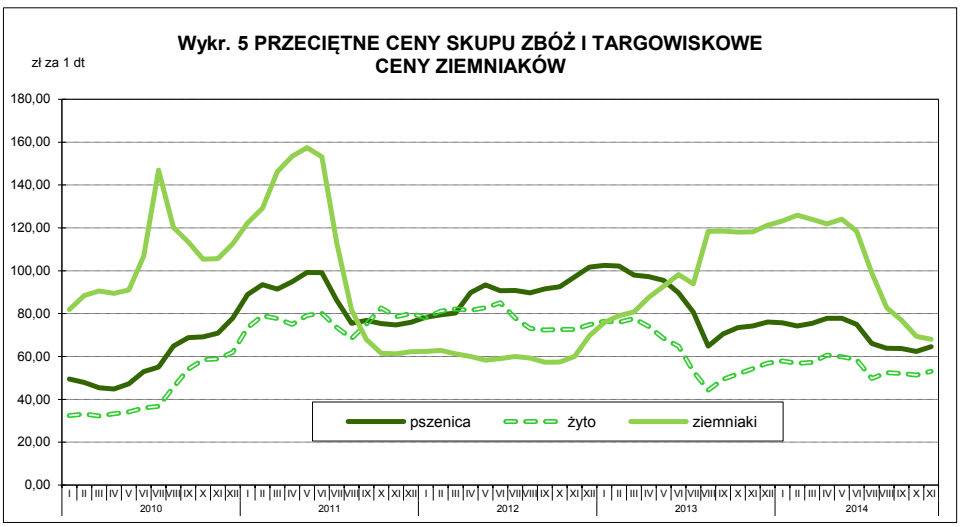
Wykr. 4 WSKAŹNIKI CEN
(analogiczny okres roku poprzedniego=100)



Na rynku rolnym w listopadzie 2014 r. ceny większości produktów pochodzenia roślinnego i zwierzęcego były niższe niż przed rokiem. Wynikowy szacunek przeprowadzony w listopadzie 2014 r. potwierdza, że w 2014 r. zbiory zbóż ogółem były większe od 2013 r. i w porównaniu ze średnią z lat 2006—2010. Wyższe niż w 2013 r. i od średnich z wielolecia były również zbiory rzepaku i rzepiku, buraków cukrowych, warzyw gruntowych i owoców z drzew. Zbiory ziemniaków był większe niż przed rokiem, ale mniejsze od średniej z wielolecia (wykr. 5). Zbiory owoców jagodowych były niższe niż przed rokiem, ale wyższe od średniej z lat 2006—2010.

Wzrost obrotów towarowych handlu zagranicznego (wyrażonych w złotych) w okresie styczeń—październik 2014 r. był wolniejszy niż w I półroczu 2014 r. Eksport zwiększył się w podobnej skali, jak import. Ujemne saldo obrotów pogłębiło się w porównaniu z notowanym przed rokiem. Wzrosła wartość wymiany z krajami rozwiniętymi, w tym krajami UE oraz z krajami rozwijającymi się, natomiast obroty z krajami Europy Środkowo-Wschodniej — zwłaszcza eksport —

były niższe niż przed rokiem. Wskaźnik terms of trade w okresie styczeń—wrzesień 2014 r. kształtował się korzystniej niż przed rokiem (104,2, wobec 101,9), na co wpłynął spadek cen towarów importowanych.



Dochody budżetu państwa w okresie styczeń—listopad 2014 r. wyniosły 260,3 mld zł (tj. 93,7% kwoty założonej w ustawie budżetowej na 2014 r.), a wydatki — 285,1 mld zł (odpowiednio 87,6%). Deficyt ukształtował się na poziomie 24,8 mld zł, co stanowiło 52,1% planu.

Departament Analiz i Opracowań Zbiorczych, GUS

SPIS TREŚCI

Janusz Witkowski — Pożegnanie Pana Profesora Tadeusza Walczaka **1**

STUDIA METODOLOGICZNE

Mirosław Szreder — Zmiany w strukturze całkowitego błędu badania próbkowego **4**

Piotr Sulewski — Miary związku między cechami w tablicy trójdzielczej **13**

Szymon Wójcik — Prognozowanie inflacji w Polsce na podstawie modeli autoregresji wektorowej **28**

BADANIA I ANALIZY

Bartosz Totleben — Empiryczna analiza wpływu wybranych czynników na wzrost gospodarczy **42**

Anna Turczak, Patrycja Zwiech — Zależność między miejscem zamieszkania na wsi i w mieście a głównym źródłem utrzymania ludności **60**

STATYSTYKA MIĘDZYNARODOWA

Łukasz Matuszczak — Konkurencyjność polskiego eksportu usług **76**

INFORMACJE. PRZEGLĄDY. RECENZJE

„Historia Polski w liczbach" — unikalna seria wydawnicza statystyczno-historyczna GUS (oprac. *Jan Berger*) **94**

Ogólnopolska konferencja użytkowników oprogramowania statystycznego R (oprac. *Maciej Beręsewicz*) **97**

Wydawnictwa GUS — grudzień 2014 r. (oprac. *Justyna Gustyn*) **99**

Informacja o sytuacji społeczno-gospodarczej kraju — listopad 2014 r. (oprac. *Departament Analiz i Opracowań Zbiorczych*, GUS) **101**

CONTENTS

Janusz Witkowski — Forewell to Professor Tadeusz Walczak **1**

METHODOLOGICAL STUDIES

Mirosław Szreder — Changes in the structure of the total error of sample survey **4**

Piotr Sulewski — Measure of the relationship between the characteristics of the 3-feature array **13**

Szymon Wójcik — Forecasting inflation in Poland based on vector autoregressive models **28**

SURVEYS AND ANALYSES

Bartosz Totleben — Empirical analysis of the impact of selected factors on economic growth **42**

Anna Turczak, Patrycja Zwiech — The relationship between place of residence in rural and urban areas and the main source of livelihood of the population **60**

INTERNATIONAL STATISTICS

Łukasz Matuszczak — The competitiveness of Polish exports of services **76**

INFORMATION. REVIEWS. COMMENTS

"Polish History in Numbers" — unique series of statistical and historical publishing of the CSO (by *Jan Berger*) **94**

A users' conference on the R statistical software (by *Maciej Beręsewicz*) **97**

Publications of the CSO of Poland in December 2014 (by *Justyna Gustyn*) **99**

Information on the socio-economic situation of Poland in November 2014 (by *Aggregated Studies Department*, CSO) **101**

TABLE DES MATIÈRES

Janusz Witkowski — Mots d'adieu à Monsieur le Professeur Tadeusz Walczak **1**

ÉTUDES MÉTHODOLOGIQUES

Mirosław Szreder — Changements de la structure de l'erreur totale relative à l'enquête par sondage **4**

Piotr Sulewski — Mesures de la rélation entre les caractéristiques relatives au tableau à trois dimensions **13**

Szymon Wójcik — Projections relatives à l'inflation en Pologne sur la base des modéles autoregressifs vectoriels **28**

ÉTUDES ET ANALYSES

Bartosz Totleben — Analyse empirique d'un impact de plusieurs facteurs sur la croissance économique **42**

Anna Turczak, Patrycja Zwiech — Corrélations entre la résidence rurale/urbaine et la principale source de moyens d'existence de la population **60**

STATISTIQUES INTERNATIONALES

Łukasz Matuszczak — Compétitivité relative aux exportations polonaises des services **76**

INFORMATIONS. REVUES. COMPTE-RENDUS

"Histoire de la Pologne en chiffres" — la série unique statistique et historique du GUS (par *Jan Berger*) **94**

Conférence des utilisateurs de l'application statistique R (par *Maciej Beręsewicz*) **97**

Publications du GUS — décembre 2014 (par *Justyna Gustyn*) **99**

Information sur la situation socio-économique du pays — novembre 2014 (par *Département d'Analyses et d'Élaborations Agrégées*, GUS) **101**

СОДЕРЖАНИЕ

Януш Витковски — Прощание профессором Тадеушом Вальчаком **1**

МЕТОДОЛОГИЧЕСКИЕ ИЗУЧЕНИЯ

Мирослав Шрэдэр — Изменения в структуре полной ошибки выборочного обследования **4**

Пиотр Сулевски — Меры связи между характеристиками в тройной таблице **13**

Шимон Войцик — Прогнозирование инфляции в Польше на основе модели векторной авторегрессии **28**

ОБСЛЕДОВАНИЯ И АНАЛИЗЫ

Бартош Тотлебэн — Эмпирический анализ влияния избранных факторов на экономический рост **42**

Анна Турчак, Патриция Звех — Зависимость между местом проживания в деревне и в городе и главными источниками средств на содержание населения **60**

МЕЖДУНАРОДНАЯ СТАТИСТИКА

Лукаш Матущак — Конкурентоспособность польского экспорта услуг **76**

ИНФОРМАЦИИ. ОБЗОРЫ. РЕЦЕНЗИИ

«История Польши в числах» — уникальная публикация статистическо-исторической серии ЦСУ (разраб. *Ян Бэргэр*) **94**

Конференция пользователей статистического программного обеспечения *R* (разраб. *Мацей Беренсевич*) **97**

Публикации ЦСУ — декабрь 2014 г. (разраб. *Юстина Густын*) **99**

Информация о социально-экономическом положении страны — ноябрь 2014 г. (разраб. *Отдел анализа и сводных разработок*, ЦСУ) **101**

Do Autorów

Szanowni Państwo!

- W „Wiadomościach Statystycznych" publikowane są artykuły poświęcone teorii i praktyce statystycznej, omawiające metody i wyniki badań prowadzonych przez GUS oraz przez inne instytucje w kraju i za granicą, jak również zastosowanie informatyki w statystyce oraz zmiany w systemie zbierania i udostępniania informacji statystycznej. Zamieszczane są też materiały dotyczące zastosowania w kraju metodologicznych i klasyfikacyjnych standardów międzynarodowych oraz informacje o działalności organów statystycznych i Polskiego Towarzystwa Statystycznego, a także o rozwoju myśli statystycznej i kształceniu statystycznym.

- Artykuły proponowane do opublikowania w „Wiadomościach Statystycznych" powinny zawierać oryginalne opisy zjawisk oraz autorskie wnioski i sugestie dotyczące rozwoju badań i analiz statystycznych. Dla zwiększenia właściwego odbioru nadsyłanych tekstów Autorzy powinni wyraźnie określić cel opracowania artykułu oraz jasno przedstawić wyniki, a w przypadku prezentacji przeprowadzonych badań — opisać zastosowaną metodę i osiągnięte wyniki. Przy prezentacji nowych metod analizy konieczne jest podanie przykładów ich zastosowania w praktyce statystycznej.

- Artykuły zamieszczane w „Wiadomościach Statystycznych" powinny wyrażać opinie własne Autorów. Autorzy ponoszą odpowiedzialność za treść zgłaszanych do publikacji artykułów. W razie zastrzeżeń ze strony czytelników w sprawie tych treści Autorzy zostają zobligowani do merytorycznej odpowiedzi na łamach miesięcznika.

- Po wstępnej ocenie przez Redakcję „Wiadomości Statystycznych" tematyki artykułu pod względem zgodności z profilem czasopisma, artykuły mające charakter naukowy przekazywane są dwóm niezależnym, zewnętrznym recenzentom specjalizującym się w poszczególnych dziedzinach statystyki, którzy w swojej decyzji kierują się kryterium oryginalności i jakości opracowania, w tym treści i formy, a także potencjalnego zainteresowania czytelników. Recenzje są opracowywane na drukach zaakceptowanych przez Kolegium Redakcyjne „Wiadomości Statystycznych". Recenzenci są zobowiązani do poświadczenia (na karcie recenzji) braku konfliktu interesów z Autorem. Wybór recenzentów jest poufny.

- Lista recenzentów oceniających artykuły w danym roku jest publikowana w pierwszym numerze elektronicznej wersji czasopisma.

- Autorzy artykułów, którzy otrzymali pozytywne recenzje, wprowadzają zasugerowane przez recenzentów poprawki i dostarczają redakcji zaktualizowaną wersję opracowania. Autorzy poświadczają w piśmie uwzględnienie wszystkich poprawek. Jeśli zaistnieje różnica zdań co do zasadności proponowanych zmian, należy wyjaśnić, które poprawki zostały uwzględnione, a w przypadku ich nieuwzględnienia przedstawić motywy swojego stanowiska.

- Kontroli poprawności stosowanych przez Autorów metod statystycznych dokonują redaktorzy statystyczni.

- Decyzję o publikacji artykułu podejmuje Kolegium Redakcyjne „Wiadomości Statystycznych". Podstawą tej decyzji jest szczegółowa dyskusja poświęcona omówieniu zgłoszonych przez Autorów artykułów, w której uwzględniane są opinie przedstawione w recenzjach wraz z rekomendacją ich opublikowania.

- Redakcja „Wiadomości Statystycznych" przestrzega zasady nietolerowania przejawów nierzetelności naukowej autorów artykułów polegającej na:
  a) nieujawnianiu współautorów, mimo że wnieśli oni istotny wkład w powstanie artykułu, określanemu w języku angielskim terminem „ghostwriting";
  b) podawaniu jako współautorów osób o znikomym udziale lub niebiorących udziału w opracowaniu artykułu, określanemu w języku angielskim terminem „guest authorship".

  Stwierdzone przypadki nierzetelności naukowej w tym zakresie mogą być ujawniane.

  W celu przeciwdziałania zjawiskom „ghostwriting" i „guest authorship" należy dołączyć do przesłanego artykułu oświadczenie (wzór oświadczenia zamieszczono na stronie internetowej) dotyczące:
  a) stwierdzenia, że zgłoszony artykuł jest własnym dziełem i nie narusza praw autorskich osób trzecich,
  b) wykazania wkładu w powstanie artykułu przez poszczególnych współautorów,
  c) poinformowania, że zgłoszony artykuł nie był dotychczas publikowany i nie został złożony w innym wydawnictwie.

  Główną odpowiedzialność za rzetelność przekazanych informacji, łącznie z informacją na temat wkładu poszczególnych współautorów w powstanie artykułu, ponosi zgłaszający artykuł.

- Artykuły opublikowane są dostępne w wersji elektronicznej na stronie internetowej czasopisma.

- Wersję pierwotną czasopisma stanowi wersja elektroniczna.

Redakcja zastrzega sobie prawo dokonywania w artykułach zmian tytułów, skrótów i przeredagowania tekstu i tablic, bez naruszenia zasadniczej myśli Autora.

Informacje ogólne

- Artykuły należy dostarczać pocztą elektroniczną (lub na płycie CD). Prosimy również o przesłanie jednego egzemplarza jednostronnego wydruku tekstu na adres:

a.swiderska@stat.gov.pl lub e.grabowska@stat.gov.pl
Redakcja „Wiadomości Statystycznych"
Główny Urząd Statystyczny
al. Niepodległości 208, 00-925 Warszawa

- Konieczne jest dołączenie do artykułu skróconej informacji (streszczenia) o jego treści (ok. 10 wierszy) w języku polskim i, jeżeli jest to możliwe, także w językach angielskim i rosyjskim. Streszczenie powinno być utrzymane w formie bezosobowej i zawierać: ogólny opis przedmiotu artykułu, określenie celu badania, przyjętą metodologię badania oraz ważniejsze wnioski.
- Prosimy również o podawanie słów kluczowych, przybliżających zagadnienia w artykule.
- Pytania dotyczące przesłanego artykułu, co do jego aktualnego statusu itp., należy kierować do redakcji na adres: a.swiderska@stat.gov.pl lub e.grabowska@stat.gov.pl lub tel. 22 608-32-25.

Wymogi edytorskie wydawnictwa

Artykuł powinien mieć optymalną objętość (łącznie z wykresami, tablicami i literaturą) 10—20 stron przygotowanych zgodnie z poniższymi wytycznymi:

1. Edytor tekstu — Microsoft Word, format *.doc lub *.docx.
2. Czcionka:
   - autor — Arial, wersalik, wyrównanie do lewej, 12 pkt.,
   - tytuł opracowania — Arial, wyśrodkowany, 16 pkt.,
   - tytuły rozdziałów i podrozdziałów — Times New Roman, wyśrodkowany, kursywa, 14 pkt.,
   - tekst główny — Times New Roman, normalny, wyjustowany, 12 pkt.,
   - przypisy — Times New Roman, 10 pkt.
3. Marginesy przy formacie strony A4 — 2,5 cm z każdej strony.
4. Odstęp między wierszami półtorej linii oraz interlinia przed tytułami rozdziałów.
5. Pierwszy wiersz akapitu wcięty o 0,4 cm, enter na końcu akapitu.
6. Wyszczególnianie rozmaitych kategorii należy zacząć od kropek, a numerowanie od cyfr arabskich.
7. Strony powinny być ponumerowane automatycznie.
8. Wykresy powinny być załączone w osobnym pliku w oryginalnej formie (Excel lub Corel), tak aby można było je modyfikować przy opracowaniu edytorskim tekstu. W tekście należy zaznaczyć miejsce ich włączenia. Należy także przekazać dane, na podstawie których powstały wykresy.
9. Tablice należy zamieszczać w tekście, zgodnie z treścią artykułu. W tablicach nie należy stosować rastrów, cieniowania, pogrubiania czy też podwójnych linii itp.
10. Pod wykresami i tablicami należy podać informacje dotyczące źródła opracowania.
11. Oznaczenia literowe należy wyróżniać następująco: macierze — wersalik, proste, pogrubione (np. $\mathbf{P}$, $\mathbf{N}_{ij}$); wektory — małe litery, kursywa, pogrubione (np. $\boldsymbol{w}$, $\boldsymbol{x}_i$); pozostałe zmienne — małe litery, kursywa, bez pogrubienia (np. w, x_i).
12. Stosowane są skróty: tablica — tabl., wykres — wykr.
13. Przypisy do tekstu należy umieszczać na dole strony.
14. Przytaczane w treści artykułu pozycje literatury przedmiotu należy zamieszczać podając nazwisko autora i rok wydania publikacji według wzoru: (Kowalski, 2002). Z kolei przytaczane z podaniem stron pozycje literatury przedmiotu należy zamieszczać w przypisie dolnym według wzoru: Kowalski (2002), s. 50—58.
15. Wykaz literatury należy zamieszczać na końcu opracowania według porządku alfabetycznego według wzoru: Kowalski J. (2002), *Tytuł publikacji*, Wydawnictwo X, Warszawa (bez podawania numerów stron). Literatura powinna obejmować wyłącznie pozycje przytoczone w artykule.