

Tomasz Śmiałowski, Piotr Jałowiecki

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

SYMULACYJNE METODY TRANSFORMACJI DANYCH POMIĘDZY PODZIAŁAMI ADMINISTRACYJNYMI W POLSCE

SIMULATION METHODS FOR DATA TRANSFORMING BETWEEN ADMINISTRATIVE DIVISIONS IN POLAND

Słowa kluczowe: reforma administracyjna, budżety gospodarstw domowych, transformacja danych, metoda Monte Carlo

Key words: administration reform, households budgets, data transformation, Monte Carlo method

Abstrakt. Celem badań było opracowanie metody transformacji danych ze starego na nowy podział terytorialny Polski. Jednym z popularniejszych źródeł danych wykorzystywanych w badaniach ekonomicznych i społecznych są wyniki badań budżetów gospodarstw domowych (BGD). Jeżeli okres badawczy jest odpowiednio długi, wykorzystywanie w charakterze danych źródłowych wyników badań BGD wiąże się z koniecznością uwzględniania danych przygotowanych zgodnie z różnymi podziałami terytorialnymi (na 49 przed i 16 po 1999 r.). Dostarczane przez GUS raporty nie pozwalają na przydzielenie poszczególnych gospodarstw do konkretnego powiatu, gminy lub miejscowości, a jedynie do określonego województwa lub klasy miejscowości. Przedstawiono propozycje dwóch metod takiej transformacji danych na podstawie metody wag oraz metody Monte Carlo. Uzyskane wyniki wskazują, że zaproponowane rozwiązania umożliwiają prostą klasyfikację gospodarstw domowych według przynależności do województwa niezależnie od podziału administracyjnego. Trafność zaproponowanych metod została potwierdzona przez analizę wartości odstających.

Wstęp

Badania prowadzone przez GUS są ważnym źródłem wiedzy o warunkach życia właścicieli gospodarstw domowych oraz wymiarach różnicowania społecznego. Na ich podstawie można ocenić zmiany: poziomu konsumpcji w gospodarstwach, cen wybranych usług i towarów, poziomu i źródeł dochodów, warunków mieszkaniowych, wyposażenia miejsca zamieszkania w dobra trwałego użytku, jak również sprawdzić wpływ wielu czynników o charakterze ekonomicznym, społecznym, demograficznym, czy terytorialnym na takie zmiany. Informacje zawarte w wynikach badań BGD, umożliwią również badanie poziomu ubóstwa, poziomu minimalnego wynagrodzenia, spożycia naturalnego, poziomu oraz zmian w dochodach nominalnych i realnych, rynku konsumpcyjnego, a także mogą być pomocne w oszacowaniach obciążeń podatkowych gospodarstw lub analizach mających na celu ustalanie poziomów świadczeń społecznych. Szczegółowa metodyka, zgodnie z którą przeprowadza się badania dostępna jest w wydawanej corocznie publikacji *Budżety gospodarstw domowych* [1994-2008]. Niektóre rozwiązania metodyczne przyjęte w tych badaniach różnią się od analogicznych rozwiązań stosowanych w rachunkach narodowych, co jest przyczyną rozbieżności pomiędzy wynikami badań BGD a danymi makroekonomicznymi publikowanymi m.in. w rocznikach statystycznych wydawanych przez GUS.

W wielu badaniach o charakterze cyklicznym, prowadzonych w ujęciu terytorialnym, wykorzystujących wyniki badań BGD jako dane źródłowe, m.in. dotyczących rynku wyrobów tytoniowych [Jałowiecka i in. 2012], wyrobów alkoholowych [Zmarzłowski, Orłowski 2012], czy wyposażenia gospodarstw domowych w sprzęt elektroniczny [Śmiałowski, Jałowiecki 2012], podstawowym problemem była zmiana podziału terytorialnego Polski. Od 1999 r. na mocy ustawy z 24 lipca 1998 o wprowadzeniu zasadniczego trójstopniowego podziału terytorialnego państwa (Dz.U. nr 96, poz. 603, z późn. zm.), obowiązuje podział na 16 dużych województw (wcześniej na 49 małych). W publikowanych i udostępnianych przez GUS wynikach badań BGD brakuje szczegółowych

informacji na temat podziału terytorialnego, takich jak przynależność gospodarstwa do gminy lub powiatu. Z tego powodu w wielu przypadkach nie jest możliwa jednoznaczna identyfikacja, do którego nowego województwa trafiły poszczególne gospodarstwa domowe. W konsekwencji, wiele spośród wspomnianych badań wykorzystujących wyniki badań BGD jako dane źródłowe ogranicza się do okresu po 1999 r. Problem ten dotyczy ok. 32% gospodarstw znajdujących się w bazie BGD udostępnianej przez GUS.

Celem badań było opracowanie metody transformacji danych ze starego na nowy podział terytorialny Polski.

Material i metodyka badań

Opracowano klasyfikację migracji starych do nowych województw z uwzględnieniem 6 klas wielkości miast: powyżej 500 tys. mieszkańców (KLM6), od 200 do 500 tys. mieszkańców (KLM5), od 100 do 200 tys. mieszkańców (KLM4), od 20 do 100 tys. mieszkańców (KLM3), poniżej 20 tys. mieszkańców (KLM2) i mieszkańców wsi (KLM1). W tym celu posłużono się metodą wag, w której kierowano się strukturą miejscowości z podziałem na klasy. W pierwszym kroku zebrano najpierw informacje na temat struktury klas miejscowości w poszczególnych województwach oraz na podstawie danych na temat reformy administracyjnej przeprowadzonej w 1998 r., zidentyfikowano stare województwa, które w całości trafiły do nowych (kategoria A) oraz takie, które zostały podzielone na kilka mniejszych części i rozparcelowane pomiędzy kilka nowych województw (kategoria B). Następnie wyznaczono procentowe udziały dla każdej z rozparcelowanych części starych województw z uwzględnieniem podziału na 6 klas miejscowości. Wyznaczone udziały procentowe stały się podstawą przypisania wszystkim gospodarstwom ze starych województw należących do kategorii B odpowiednich wag definiujących prawdopodobieństwo przypisania do określonego nowego województwa. Oczywiście w przypadku starych województw należących do kategorii A, jedynym zabiegiem było przydzielenie ich do danego nowego województwa. Algorytm przygotowano jako odwracalny, tak aby konwersja była możliwa w obie strony – ze starego na nowy podział terytorialny i odwrotnie.

Zgodnie z klasyczną definicją, „metoda Monte Carlo jest metodą reprezentującą rozwiązanie problemu w postaci parametru pewnej hipotetycznej populacji i używająca sekwencji liczb losowych do skonstruowania próby losowej danej populacji, z której to statystyczne oszacowania tego parametru mogą być otrzymane” [Halton 1970]. Algorytm metody Monte Carlo jest probabilistyczny, a każda kolejna przetwarzana instrukcja definiowana jest w sposób losowy. Jego głównym zadaniem jest oszacowanie wartości oczekiwanej zmiennej losowej.

Zgodnie z założeniami metody, zmienna losowa \underline{y} o funkcji gęstości prawdopodobieństwa $f(\underline{y})$ i wartość oczekiwana zdefiniowana zgodnie z formułą (1) są niezależnymi zmiennymi losowymi o funkcji gęstości prawdopodobieństwa $f_{\underline{y}}(\underline{y})$.

$$\mu_y = E\{\underline{y}\} = \int_{-\infty}^{\infty} \underline{y} f_{\underline{y}}(\underline{y}) d\underline{y}, \text{ gdzie } \{\underline{y}_n\} = \{\underline{y}_n | n = 1, 2, \dots, N\} \quad (1)$$

Ich średnia arytmetyczna (2) jest również zmienną losową, która jest estymatorem wartości oczekiwanej zmiennej losowej \underline{y} .

$$\hat{\mu}_{\underline{y}} = \frac{1}{N} \sum_{n=1}^N \underline{y}_n \quad (2)$$

Według założeń centralnego twierdzenia granicznego przy $N \rightarrow \infty$ estymator $\hat{\mu}_{\underline{y}}$ dąży do zmiennej losowej o rozkładzie normalnym z wartością oczekiwaną μ_y i wariancją (3).

$$\text{Var}\{\hat{\mu}_{\underline{y}}\} = \frac{1}{N^2} \sum_{n=1}^N \text{Var}\{\underline{y}_n\} = \frac{\sigma_y^2}{N} \quad (3)$$

Skuteczne stosowanie metody Monte Carlo wymaga wykorzystania dużej ilości liczb losowych (zwykle kilka tysięcy), które pozyskiwane są za pomocą generatora liczb pseudolosowych. Dodatkowo muszą one charakteryzować się podobieństwem cech podstawowych w stosunku do odpowiednich liczb losowych, cechujących się rozkładem równie prawdopodobnym w przedziale domkniętym $\langle 0, 1 \rangle$.

Weryfikację wyników przeprowadzonych transformacji przeprowadzono na podstawie oceny występowania wartości odstających (ang. *outliers*) typu rampowego LS (ang. *level shift*) dla sześćdziesiątej pierwszej obserwacji odpowiadającej styczniowi 1999 r. Był to punkt czasowy, w którym zmienił się podział terytorialny Polski. Wartość odstająca typu LS oznacza znaczącą skokową zmianę wartości w szeregu czasowym mającą charakter trwały. Wszystkie wartości w badanych szeregach czasowych testowano pod kątem istnienia wartości odstającej w sposób iteracyjny, wykorzystując w tym celu model regARIMA oraz statystykę testową (4).

$$RS_i = \frac{y_i - \hat{y}_i}{\sqrt{MSE_i(1-\alpha)}} \quad (4)$$

gdzie:

y_i – wartość i -tej obserwacji w rzeczywistym szeregu czasowym,

\hat{y}_i – wartość i -tej obserwacji uzyskana w wyniku zastosowania odpowiedniego modelu regARIMA,

MSE_i – średni błąd kwadratowy bez uwzględnienia w zastosowanym modelu regARIMA i -tej obserwacji,

α – poziom istotności.

W przypadku, gdy wartość RS_i dla i -tej obserwacji była większa niż wartość krytyczna, wynosząca dla szeregów czasowych składających się z 180 obserwacji 3,95, uznawano ją za potencjalnie odstającą [Weisenberg 1985, Ljung 1993]. Następnie przeprowadzano automatyczną identyfikację z wykorzystaniem algorytmu detekcji wartości odstających wbudowanego w procedurę ARIMA-X12, która potwierdzała lub nie taki charakter testowanej obserwacji w badanym szeregu [Bell 1983, Chang i in. 1988, Otto, Bell 1990]. Wszystkie obliczenia wykonano w pakiecie SAS Enterprise z wykorzystaniem programów opartych na procedurach ARIMA i X12.

Wyniki badań

Od 1 stycznia 1999 r. zaczęła w Polsce obowiązywać reforma podziału administracyjnego, na mocy której wprowadzono 3-stopniową strukturę podziału terytorialnego oraz przywrócono zlikwidowane w 1975 r. powiaty. W wyniku reformy zredukowano liczbę województw z 49 do 16, a w miastach, które utraciły prawa miast wojewódzkich, utworzono grupę powiatów grodzkich (miast na prawach powiatu).

Stare województwa podzielono na dwie kategorie: A – które w całości trafiły do jednego nowego województwa oraz B – które rozparcelowano pomiędzy kilka nowych województw. Do kategorii A zaliczono 22 stare województwa, które weszły w całości w skład 13 nowych województw:

- dolnośląskiego: wałbrzyskie, wrocławskie, jeleniogórskie, legnickie;
- kujawsko-pomorskiego: włocławskie;
- lubelskiego: lubelskie stare, zamojskie, chełmskie;
- łódzkiego: łódzkie stare, sieradzkie;
- małopolskiego: krakowskie, nowosądeckie;
- mazowieckiego: warszawskie;
- opolskiego: opolskie stare;
- podkarpackiego: rzeszowskie, przemyskie;
- podlaskiego: białostockie;
- pomorskiego: gdańskie;
- warmińsko-mazurskiego: olsztyńskie;
- wielkopolskiego: poznańskie;
- zachodniopomorskiego: szczecińskie, koszalińskie.

W tabeli 1 przedstawiono klasyfikację starych województw należących do kategorii B, obejmującą ich migracje do nowych województw z uwzględnieniem 6 klas miejscowości.

Tabela 1. Udział klasy miejscowości w nowym podziale
 Table 1. The share class of the new division

| Stare województwo/ Old province | Nowe województwo/ New province | Klasy wielkości/Class size [%] | | | | | |
|------------------------------------|-----------------------------------|--------------------------------|--------|--------|--------|--------|------|
| | | KLM1 | KLM2 | KLM3 | KLM4 | KLM5 | KLM6 |
| Białkopodlaskie | lubelskie | 78,95 | 80,00 | 100,00 | 0,00 | 0,00 | 0,00 |
| | mazowieckie | 21,05 | 20,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Bielskie | małopolskie | 65,64 | 46,15 | 50,00 | 0,00 | 0,00 | 0,00 |
| | śląskie | 34,36 | 53,85 | 50,00 | 100,00 | 0,00 | 0,00 |
| Bydgoskie | kujawsko-pomorskie | 77,94 | 86,96 | 75,00 | 0,00 | 100,00 | 0,00 |
| | pomorskie | 18,65 | 8,70 | 25,00 | 0,00 | 0,00 | 0,00 |
| | wielkopolskie | 3,41 | 4,35 | 0,00 | 0,00 | 0,00 | 0,00 |
| Ciechanowskie | warmińsko-mazurskie | 10,32 | 14,29 | 25,00 | 0,00 | 0,00 | 0,00 |
| | mazowieckie | 89,68 | 85,71 | 75,00 | 0,00 | 0,00 | 0,00 |
| Częstochowskie | łódzkie | 19,97 | 7,14 | 0,00 | 0,00 | 0,00 | 0,00 |
| | opolskie | 19,72 | 28,57 | 0,00 | 0,00 | 0,00 | 0,00 |
| | śląskie | 47,50 | 57,14 | 100,00 | 0,00 | 100,00 | 0,00 |
| | świętokrzyskie | 12,80 | 7,14 | 0,00 | 0,00 | 0,00 | 0,00 |
| Elbląskie | warmińsko-mazurskie | 40,00 | 60,00 | 0,00 | 100,00 | 0,00 | 0,00 |
| | pomorskie | 60,00 | 40,00 | 100,00 | 0,00 | 0,00 | 0,00 |
| Gorzowskie | lubuskie | 26,37 | 60,00 | 100,00 | 100,00 | 0,00 | 0,00 |
| | wielkopolskie | 11,28 | 5,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| | zachodniopomorskie | 62,35 | 35,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Kaliskie | dolnośląskie | 4,20 | 12,50 | 0,00 | 0,00 | 0,00 | 0,00 |
| | łódzkie | 15,91 | 6,25 | 0,00 | 0,00 | 0,00 | 0,00 |
| | wielkopolskie | 79,89 | 81,25 | 100,00 | 100,00 | 0,00 | 0,00 |
| Katowickie | małopolskie | 25,21 | 13,64 | 4,35 | 0,00 | 0,00 | 0,00 |
| | śląskie | 74,79 | 86,36 | 95,65 | 100,00 | 100,00 | 0,00 |
| Kieleckie | małopolskie | 9,25 | 5,88 | 0,00 | 0,00 | 0,00 | 0,00 |
| | świętokrzyskie | 90,75 | 94,12 | 100,00 | 0,00 | 100,00 | 0,00 |
| Konińskie | łódzkie | 12,94 | 6,67 | 0,00 | 0,00 | 0,00 | 0,00 |
| | wielkopolskie | 87,06 | 93,33 | 100,00 | 0,00 | 0,00 | 0,00 |
| Krośnińskie | małopolskie | 3,34 | 11,11 | 0,00 | 0,00 | 0,00 | 0,00 |
| | podkarpackie | 96,66 | 88,89 | 100,00 | 0,00 | 0,00 | 0,00 |
| Leszczyńskie | dolnośląskie | 19,25 | 12,50 | 0,00 | 0,00 | 0,00 | 0,00 |
| | lubuskie | 4,94 | 12,50 | 0,00 | 0,00 | 0,00 | 0,00 |
| | wielkopolskie | 75,81 | 75,00 | 100,00 | 0,00 | 0,00 | 0,00 |
| Łomżyńskie | podlaskie | 89,00 | 100,00 | 100,00 | 0,00 | 0,00 | 0,00 |
| | mazowieckie | 11,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Ostrołęckie | warmińsko-mazurskie | 2,42 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| | mazowieckie | 97,58 | 100,00 | 100,00 | 0,00 | 0,00 | 0,00 |
| Piłskie | wielkopolskie | 80,74 | 85,00 | 75,00 | 0,00 | 0,00 | 0,00 |
| | zachodniopomorskie | 19,26 | 15,00 | 25,00 | 0,00 | 0,00 | 0,00 |
| Piotrkowskie | świętokrzyskie | 5,34 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| | łódzkie | 94,66 | 100,00 | 100,00 | 0,00 | 0,00 | 0,00 |
| Płockie | mazowieckie | 58,84 | 57,14 | 50,00 | 100,00 | 0,00 | 0,00 |
| | łódzkie | 41,16 | 42,86 | 50,00 | 0,00 | 0,00 | 0,00 |

Tabela 1. Cd./ Table 1. Cont.

| | | | | | | | |
|----------------|---------------------|-------|-------|--------|--------|--------|------|
| Radomskie | łódzkie | 1,18 | 7,69 | 0,00 | 0,00 | 0,00 | 0,00 |
| | mazowieckie | 97,58 | 92,31 | 100,00 | 0,00 | 100,00 | 0,00 |
| | świętokrzyskie | 1,24 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Siedleckie | mazowieckie | 79,37 | 90,00 | 66,67 | 0,00 | 0,00 | 0,00 |
| | lubelskie | 20,63 | 10,00 | 33,33 | 0,00 | 0,00 | 0,00 |
| Skierniewickie | mazowieckie | 50,51 | 25,00 | 50,00 | 0,00 | 0,00 | 0,00 |
| | łódzkie | 49,49 | 75,00 | 50,00 | 0,00 | 0,00 | 0,00 |
| Słupskie | zachodniopomorskie | 10,51 | 11,11 | 0,00 | 0,00 | 0,00 | 0,00 |
| | pomorskie | 89,49 | 88,89 | 100,00 | 100,00 | 0,00 | 0,00 |
| Suwalskie | warmińsko-mazurskie | 44,36 | 81,82 | 50,00 | 0,00 | 0,00 | 0,00 |
| | podlaskie | 55,64 | 18,18 | 50,00 | 0,00 | 0,00 | 0,00 |
| Tarnobrzeskie | lubelskie | 23,64 | 14,29 | 33,33 | 0,00 | 0,00 | 0,00 |
| | podkarpackie | 5,66 | 28,57 | 33,33 | 0,00 | 0,00 | 0,00 |
| | świętokrzyskie | 70,70 | 57,14 | 33,33 | 0,00 | 0,00 | 0,00 |
| Tarnowskie | małopolskie | 78,20 | 75,00 | 50,00 | 100,00 | 0,00 | 0,00 |
| | podkarpackie | 21,80 | 25,00 | 50,00 | 0,00 | 0,00 | 0,00 |
| Toruńskie | kujawsko-pomorskie | 85,95 | 88,89 | 100,00 | 100,00 | 100,00 | 0,00 |
| | warmińsko-mazurskie | 14,05 | 11,11 | 0,00 | 0,00 | 0,00 | 0,00 |
| Zielonogórskie | lubuskie | 87,95 | 91,30 | 100,00 | 100,00 | 0,00 | 0,00 |
| | wielkopolskie | 12,05 | 8,70 | 0,00 | 0,00 | 0,00 | 0,00 |

Źródło: opracowanie własne

Source: own study

Po dokonaniu klasyfikacji wszystkich gospodarstw domowych z uwzględnieniem migracji pomiędzy województwami w latach 1994-1998, zgodnie z metodą, przygotowano 16 miesięcznych szeregów czasowych zawierających przeciętną wysokość konsumpcji wódki w gospodarstwie domowym dla każdego województwa obejmujących lata 1994-2008. Analogicznie postąpiono wykorzystując jako metodę klasyfikacji metodę stochastyczną Monte Carlo. Następnie dokonano identyfikacji wartości odstających w przygotowanych szeregach (zidentyfikowany *outlier* oznaczony został znakiem „+”). Ich zestawienie dla metody wag przedstawiono w tabeli 2, a dla metody Monte Carlo w tabeli 3.

Tabela 2. Wyniki identyfikacji wartości odstających dla szeregów wojewódzkich wyznaczonych zgodnie z metodą wag

Table 2. The results of identification of outliers for province's time series appointed with using of weights method

| Województwa/Provinces | Szereg czasowy/Time serie | | | | | | | | | |
|--|---------------------------|---|---|---|---|----|----|----|----|--|
| | 2 | 3 | 4 | 5 | 9 | 10 | 11 | 12 | 13 | |
| Dolnośląskie, wielkopolskie | + | | + | | | | | + | + | |
| Kujawsko-pomorskie, małopolskie, śląskie | | + | | | | | | + | + | |
| Lubelskie, mazowieckie | | + | | + | | | | + | + | |
| Lubuskie | | + | | | | | + | + | + | |
| Łódzkie, podkarpackie, świętokrzyskie | | + | + | | | | | + | + | |
| Opolskie | + | + | | | | | | + | + | |
| Podlaskie | | | + | + | | | | + | + | |
| Pomorskie | | | + | | + | | | + | + | |
| Warmińsko-Mazurskie | | + | | | | + | | + | + | |
| Zachodniopomorskie | | | + | | | | + | + | + | |

Źródło: opracowanie własne

Source: own study

Tabela 3. Wyniki identyfikacji wartości odstających dla szeregów wojewódzkich wyznaczonych zgodnie z metodą Monte Carlo

Table 3. The results of identification of outliers for province time series appointed with using of Monte Carlo method

| Województwa/Provinces | Szereg czasowy/Time serie | | | | | | | | | |
|---|---------------------------|---|---|---|---|----|----|----|----|--|
| | 2 | 3 | 4 | 5 | 6 | 10 | 11 | 12 | 13 | |
| Dolnośląskie, wielkopolskie | + | | + | | | | | + | + | |
| Kujawsko-pomorskie, małopolskie, opolskie | | + | | | | | + | + | + | |
| Lubelskie, mazowieckie | | + | | + | | | | + | + | |
| Lubuskie, łódzkie, świętokrzyskie | | + | + | | | | | + | + | |
| Podkarpackie | | | + | | | + | | + | + | |
| Podlaskie, warmińsko-mazurskie | | + | | | + | | | + | + | |
| Pomorskie | + | | | | | + | | + | + | |
| Śląskie | | + | | | | | | + | + | |
| Zachodniopomorskie | | | + | | | | + | + | + | |

Źródło: opracowanie własne

Source: own study

W żadnym z badanych szeregów czasowych nie stwierdzono istnienia wartości odstającej w przypadku 61 obserwacji, czyli w punkcie czasowym odpowiadającym styczniowi 1999 r. Oznacza to, że zaprezentowane metody transformacji danych pomiędzy starym i nowym podziałem terytorialnym umożliwiły skuteczną integrację dwóch szeregów czasowych: dla lat 1994-1998 i 1999-2008. Należy przy tym podkreślić, że w wyniku zastosowania obydwu metod uzyskano bardzo zbliżone wartości przeciętnej konsumpcji wódki we wszystkich województwach. Stwierdzone wartości odstające dotyczyły pierwszych 13 obserwacji. W przypadku wszystkich badanych województw, jako wartość odstającą sklasyfikowano 13 obserwację, która odpowiadała styczniowi 1995 r. Wyjaśnieniem tej regularności jest fakt, że w 1994 r., czyli w ciągu pierwszych 12 miesięcy badanych szeregów czasowych dane na temat konsumpcji wódki były zbierane jedynie w postaci wartościowej, natomiast później również w postaci ilościowej. W związku z tym wystąpiła konieczność przeliczenia danych wartościowych dla 1994 r. na ilościowe z wykorzystaniem średniej ceny wódki w poszczególnych województwach. Z reguły taka „systemowa” zmiana sposobu jest przyczyną wystąpienia wartości odstających typu LS.

Podsumowanie

Przedstawione metody pozwalają na klasyfikowanie w prosty sposób gospodarstw domowych według przynależności do województwa bez konieczności stosowania dwóch różnych podziałów terytorialnych dla okresów przed i po 1999 r. Skuteczność badanych metod potwierdzono badając występowanie wartości odstających w miesięcznych szeregach czasowych zawierających przeciętny poziom konsumpcji wódki dla każdego województwa osobno. Łącznie zbadano 32 szeregi czasowe i w żadnym nie stwierdzono wartości odstających, które mogłyby sugerować nieskuteczność zaproponowanych metod –wag i Monte Carlo. Procedury transformacji z wykorzystaniem metody Monte Carlo oraz wyznaczania wartości odstających charakteryzują się dużym stopniem komplikacji, jednak można je stosunkowo łatwo zaimplementować w większości popularnych pakietów analityczno-statystycznych. Opisane metody zastosowano z powodzeniem w badaniach terytorialnego zróżnicowania obszarów wykluczeń technologicznych, w których zostały wykorzystane do transformacji danych dotyczących m.in. posiadania komputera, telefonu komórkowego i połączenia do internetu [Śmiałowski, Jałowiecki 2012].

Literatura

- Bell W.R. 1983: *A Computer Program for Deceting Outliers in Time Series*. Proceedings of the American Statistical Association, Business and Economic Statistic Section, s. 634-639.
- Budżety gospodarstw domowych w 2011 roku*. 2012: GUS, Warszawa.
- Chang I., Tiao G.C., Chen C. 1988: *Estimation of Time Series Parameters in the Presence of Outliers*. Technometrics, vol. 20, no. 2, s. 193-204.
- Demidovič B.P., Maron A. 1965: *Metody numeryczne, analiza, algebra, metody Monte Carlo*, PWN, Warszawa.
- Halton J.H. 1970: *A retrospective and prospective survey of the Monte Carlo method*. SIAM Review, 12, s. 1-63.
- Jałowiecka E., Jałowiecki P., Orłowski A. 2012: *Demograficzne zróżnicowanie konsumpcji papierosów w Polsce w latach 1999-2008*. Roczn. Nauk. SERiA, t. XIV, z. 4, s. 40-44.
- Ljung G.M. 1993: *An Outlier Detection in Time Series*. Journal of the Royal Statistical Society, B, vol. 55, p. 559-567.
- Otto M.C., Bell W.R. 1990: *Two Issues in Time Series Outlier Detection Using Indicator Variables*, Proceedings of the American Statistical Association, Business and Economic Statistics Section, p. 182-187.
- Patrykiewicz A. 1998: *Wprowadzenie do metody Monte Carl*, Wyd. UMCS, Lublin.
- Śmiałowski T., Jałowiecki P. 2012: *Terytorialne zróżnicowanie obszarów wykluczeń technologicznych*, Roczn. Nauk. SERiA, s. 125-130.
- Weisenberg S. 1985: *Applied Linear Regression*. 2nd Edition, John Wiley and Sons, New York, NY, USA.
- Zmarzłowski K., Orłowski A. 2012: *Zmiany na rynku wyrobów spirytusowych w latach 1999-2009*, Roczn. Nauk. SERiA, s. 599-604.

Summary

One of the most popular data sources used in economic and social studies are the results of households budgets survey (BGD). If the time horizon of study is long enough, the using of BGD results as a data source is connected with take account necessity of data prepared in different administration divisions (on 49 and 16 voivodeships after 1999). Reports delivered by GUS do not allow for explicit allocation of households to a particular county, municipality or town, but only for voivodeship and town or village class. In paper, proposition of two methods such data transformation with using of weights and Monte Carlo simulations are presented. The results show that the proposed solutions provide a simple classification of households by belonging to the voivodeship regardless of the administrative division. Accuracy of the proposed method is confirmed by the analysis of outliers.

Adres do korespondencji
mgr inż. Tomasz Śmiałowski
Szkoła Główna Gospodarstwa Wiejskiego
Wydział Zastosowań Informatyki i Matematyki
ul. Nowoursynowska 166
02-787 Warszawa
e-mail: tomasz_smialowski@sggw.pl