

Wykorzystanie metod statystyki małych obszarów do tworzenia map ubóstwa w Polsce

W ostatnich latach obserwuje się w Polsce zwiększony popyt na szczegółowe informacje o zjawiskach lokalnych. Niestety, badania statystyki publicznej są najczęściej reprezentatywne dla kraju lub województw. Na niższych poziomach agregacji przestrzennej (podregionów czy powiatów) mała liczebność próby powoduje, że takie szacunki cechują się nieakceptowaną precyzją. W takich przypadkach stosuje się metody statystyki małych obszarów, które wykorzystują wszystkie dostępne informacje (Dehnel, 2010; Gołata, 2004). Uzyskane rezultaty charakteryzują się wówczas akceptowalną precyzją bez konieczności zwiększania liczebności próby, a co za tym idzie kosztów badania.

Jednym z głównych celów Banku Światowego — zgodnie z milenijnymi celami rozwoju — jest eliminacja ubóstwa i głodu na świecie. Mapy ubóstwa są tu bardzo pomocne. Dostarczają one szczegółowych informacji o terytorialnym zróżnicowaniu ubóstwa i wspierają procesy decyzyjne przy rozdzielaniu funduszy w programach polityki spójności (Bedi i in., 2007).

Projekt *Poverty Mapping in the new EU Member States (Mapy ubóstwa w nowych krajach członkowskich Unii Europejskiej)* ma na celu opracowanie map ubóstwa w 11 krajach, które weszły do Unii Europejskiej w 2004 r. i później. Jego głównym założeniem jest wykorzystanie Europejskiego Badania Dochodów i Warunków Życia (EU-SILC) oraz danych ze spisu powszechnego w celu uzyskania szacunków wskaźnika zagrożenia ubóstwem na poziomie NUTS 3 lub niższym. Na mocy podpisanego listu intencyjnego, celem współpracy Departamentu Badań Społecznych Warunków Życia w GUS i Ośrodka Statystyki Małych Obszarów Urzędu Statystycznego w Poznaniu z Bankiem Światowym było wykorzystanie techniki stosowanej w statystyce małych obszarów do stworzenia map ubóstwa na poziomie podregionów (NUTS 3).

W artykule opisano model wynikowy powstały w czasie pracy nad mapami ubóstwa w ramach wspomnianego projektu (*Mapy ubóstwa...*, 2014). Przedstawione wyniki mają charakter eksperymentalny i nie są oficjalnymi danymi GUS.

OBECNY STAN BADAŃ

W badaniu EU-SILC przyjęto ekonomiczną definicję ubóstwa rozpatrywaną w kategorii dochodów. Osoby w gospodarstwie domowym uznawane są za ubogie, gdy ich dochód do dyspozycji (po uwzględnieniu transferów społecznych) jest niższy od granicy ubóstwa ustalonej jako 60% krajowej mediany dochodów ekwiwalentnych do dyspozycji (*Dochody...*, 2012). Odsetek osób w ubogich

gospodarstwach domowych w relacji do wszystkich osób nazywany jest wskaźnikiem zagrożenia ubóstwem. W 2011 r. poniżej tej granicy znajdowało się w Polsce 17,7% osób w gospodarstwach domowych.

Na podstawie EU-SILC szacowany jest także wskaźnik zagrożenia ubóstwem w przekroju regionów i województw. Na wyk. 1 przedstawiono oszacowanie wskaźnika zagrożenia ubóstwem na poziomie województw.

Według badania EU-SILC w roku 2011 w Polsce wskaźnikiem zagrożenia ubóstwem wyższym od wartości dla całego kraju cechowało się 8 województw. Najgorszą sytuację odnotowano w województwach położonych na południowym wschodzie kraju. Wartość wskaźnika przekracza tam 20%, co oznacza, że w co piątym gospodarstwie żyją osoby zagrożone niedostatkiem. Woj. lubelskie osiągnęło najwyższą wartość ze wszystkich obszarów Polski — 31,3%. Duże zagrożenie niedostatkiem wystąpiło w województwach świętokrzyskim (25,5%) i lubuskim (25,1%). Z kolei najniższe wartości stopy ubóstwa odnotowano w województwach dolnośląskim (12,8%), śląskim (13%) oraz opolskim (13,4%).

Należy pamiętać, że przedstawione wyniki oszacowano na podstawie badania reprezentacyjnego, zatem charakteryzują się one występowaniem błędów losowych. Rozmiar takiego błędu rośnie wraz ze spadkiem wielkości próby dla danego obszaru. W 2011 r. w badaniu EU-SILC zbadano 28305 osób mieszkających w 12871 gospodarstwach domowych. W ujęciu wojewódzkim liczebność ta wahała się w przedziale 808—3934 osób. Konsekwencją mniejszej próby jest

wzrost wariancji otrzymanych oszacowań. Oceniając jakość estymacji można posłużyć się wartością wskaźnika precyzji wyrażonego współczynnikiem zmienności (*coefficient of variation*):

$$CV = \frac{\sqrt{Var(\hat{Y})}}{\hat{Y}} \cdot 100 \quad (1)$$

gdzie:

$\hat{Y}$ — estymator wartości globalnej,
 $Var(\hat{Y})$ — oszacowanie wariancji estymatora.

Przyjmuje się, że oszacowanie, dla którego CV przekracza 10% należy interpretować z ostrożnością, natomiast w przypadkach, gdy jest większe niż 20%, to wyniki estymacji nie powinny być publikowane (*Ludność...*, 2013).

W przedstawionej estymacji wskaźnika zagrożenia ubóstwem dla województw poszczególne jednostki charakteryzowały się współczynnikiem zmienności z przedziału 7,0—19,6%. W przypadku województw podlaskiego, opolskiego i zachodniopomorskiego do otrzymanych szacunków należy odnosić się z dystansem, ponieważ cechują się dość szerokim przedziałem ufności. Przykładowo, dla woj. podlaskiego można stwierdzić, że z przyjętym poziomem ufności równym 0,95, przedział wartości od 10,0% do 22,4% pokrywa rzeczywistą wartość wskaźnika zagrożenia ubóstwem.

Mając na uwadze wymienione kwestie podjęto próbę estymacji wskaźnika zagrożenia ubóstwem w podregionach. W tym celu wykorzystano metodę statystyki małych obszarów, a konkretnie model Faya-Herriota oraz zbudowany na jego podstawie estymator *EBLUP*.

MODEL FAYA-HERRIOTA W STATYSTYCE MAŁYCH OBSZARÓW

Model Faya-Herriota jest najczęściej wykorzystywanym modelem w przypadku, kiedy dostępne są jedynie dane dla danego obszaru (Pfeffermann, 2013). Jest to model z klasy liniowych modeli mieszanych i został zaproponowany w 1979 r. (Fay, Herriot, 1979). Może on być zapisany następująco:

$$\hat{\theta}_{y,d} = \theta_{x,d}^T \beta + v_d + \varepsilon_d \quad (2)$$

gdzie:

$\hat{\theta}_{y,d}$ — estymator bezpośredni wartości globalnej dla cechy y w obszarze d ,
 $\theta_{x,d}$ — wektor cech objaśniających dla obszaru d ,

β — wektor współczynników,
 v_d — efekt obszaru d ,
 $v_d \sim N(0, \sigma_v^2) \varepsilon_d$ — błąd losowy dla obszaru d , $\varepsilon_d \sim N(0, \sigma_d^2)$.

NAJLEPSZY EMPIRYCZNY NIEOBCIĄŻONY ESTYMATOR LINIOWY

Najlepszy empiryczny nieobciążony estymator liniowy (*Empirical Best Linear Unbiased Predictor* — *EBLUP*) modelu Faya-Herriota przyjmuje następującą postać:

$$\hat{\theta}_{y,d}^{eblop} = \gamma_d \hat{\theta}_{y,d} + (1 - \gamma_d) \theta_{x,d}^T \hat{\beta} \quad (3)$$

gdzie:

$\hat{\theta}_{y,d}^{eblop}$ — estymator *EBLUP* dla cechy y w obszarze d ,
 $\hat{\theta}_{y,d}$ — estymator bezpośredni wartości globalnej dla cechy y w obszarze d ,
 $\theta_{x,d}^T \hat{\beta}$ — estymator syntetyczny wyznaczony na podstawie modelu w obszarze d ,
 γ_d — wagą, która wyrażona jest wzorem:

$$\gamma_d = \frac{\hat{\sigma}_v^2}{\sigma_d^2 + \hat{\sigma}_v^2} \quad (4)$$

gdzie:

$\hat{\sigma}_v^2$ — wariancja wynikająca z modelu estymowana metodą momentów,
 σ_d^2 — wariancja estymatora bezpośredniego wynikająca ze schematu losowania próby.

EBLUP jest liniową, ważoną kombinacją estymatora bezpośredniego oraz estymatora syntetycznego. Waga γ_d przyjmuje wartości z przedziału 0—1 i jest przypisana do każdego analizowanego obszaru d . Wielkość ta jest zależna od wartości wariancji dla estymatora bezpośredniego. Większa waga jest przypisywana do estymatora bezpośredniego w przypadku, gdy jego wariancja jest mała. Innymi słowy, estymator *EBLUP* przyjmuje wartość estymatora bezpośredniego, kiedy cechuje się odpowiednią precyzją, natomiast w przeciwnym przypadku — wartość estymatora syntetycznego (Boonstra, Buelens, 2011).

DANE

W analizie wykorzystano dane pochodzące z EU-SILC 2011, Narodowego Spisu Ludności i Mieszkań 2011 oraz Banku Danych Lokalnych. Na podstawie badania EU-SILC oszacowano wskaźnik zagrożenia ubóstwem w przekroju podregionów, który następnie posłużył jako zmienna objaśniana. Mimo bogactwa badanej charakterystyki wykorzystanie innych zmiennych z tego badania w charakterze cech objaśniających doprowadziłoby do zwiększenia błędu losowego i obciążenia oszacowań parametrów β modelu. Na skutek tego w charakterze zmiennych objaśniających rozważano dane pochodzące z pozostałych dwóch źródeł. Wyróżniono następujące dziedziny, w których definiowano potencjalne determinanty ubóstwa:

- demografia (udział różnych grup wieku, płci, poziomu wykształcenia, statusu cywilnego w populacji);
- status na rynku pracy (udział aktywnych, pracujących i bezrobotnych w odpowiedniej grupie);
- infrastruktura mieszkaniowa (powierzchnia mieszkania na osobę, dostęp do: elektryczności, kanalizacji, centralnego ogrzewania, gazu, prysznica lub wanny);
- charakterystyka gospodarstwa domowego (liczba członków gospodarstwa domowego i pokoi przypadających na osobę, wykształcenie członków gospodarstwa domowego);
- ochrona zdrowia i opieka społeczna;
- podział terytorialny.

Wszystkie analizowane zmienne mierzono na skali ilościowej i tam, gdzie to było uzasadnione nadawano im względny charakter. Łącznie analizowano i brano pod uwagę ponad 200 potencjalnych cech objaśniających.

MODEL

Wybór zmiennych do modelu był podyktowany przesłankami merytorycznymi. Starano się ustalić zależność przyczynowo-skutkową pomiędzy analizowanymi cechami i wskaźnikiem zagrożenia ubóstwem. W niektórych przypadkach nie było to oczywiste, ponieważ trudno stwierdzić, czy poziom analizowanej cechy był spowodowany wysokim poziomem ubóstwa w województwie czy też duże ubóstwo przyczyniło się do wystąpienia takiego poziomu danego zjawiska. Przykładem takiej cechy rozpatrywanej w kategorii gospodarstwa jest wykształcenie. Mianowicie, bez szczegółowej analizy jednostki nie można określić, czy ubóstwo zostało spowodowane przez niskie wykształcenie czy też bieda panująca w gospodarstwie domowym nie pozwoliła na zdobycie wyższego wykształcenia (Haughton, Khandker, 2009).

Dopiero po zidentyfikowaniu potencjalnych zmiennych objaśniających sprawdzano spełnienie warunków statystycznych, czyli istotność zmiennej po

dodaniu jej do modelu, znak współczynnika przy parametrze β oraz wpływ cechy na wielkość współczynnika determinacji R^2 .

W rezultacie opracowano model, którego parametry zamieszczono w tabl. 1.

TABL. 1. WSPÓŁCZYNNIKI REGRESJI DOPASOWANEGO MODELU OPISUJĄCEGO WIELKOŚĆ WSKAŹNIKA ZAGROŻENIA UBÓSTWEM W PODREGIONACH

Zmienne objaśniające	Ocena parametru	Odchylenie standardowe	p-wartość
Udział osób samotnych (powyżej 25 roku życia) w populacji	1,3958	0,5209	0,0095
Udział gospodarstw domowych z dwiema osobami z wykształceniem co najwyżej zawodowym	0,3031	0,1903	0,1166
Iloraz liczby osób wymeldowanych do zameldowanych na pobyt stały	0,0199	0,0327	0,5458
Zmienna zero-jedynkowa ^a	0,0187	0,0153	0,2285
Liczba pokoi w przeliczeniu na osobę	-0,1465	0,0768	0,0614
Udział gospodarstw domowych z wanną lub prysznicem w ogólnej liczbie gospodarstw domowych	-0,7854	0,1606	0,0000
Stała	0,7437	0,2239	0,0015

^a Zmienna przyjmująca wartość równą 1, gdy gęstość zaludnienia w podregionie jest mniejsza od 33 percentyla wartości gęstości zaludnienia we wszystkich podregionach oraz 0 w przeciwnym razie.

Źródło: opracowanie własne na podstawie EU-SILC 2011 i NSP 2011.

Analizowany model w 60% wyjaśniał zmienność wskaźnika zagrożenia ubóstwem w podregionach. Nie wszystkie zmienne objaśniające cechowały się istotnością na akceptowalnym poziomie co najmniej 0,1, jednak ze względów merytorycznych oraz po konsultacji z ekspertami w dziedzinie postanowiono pozostawić je w modelu. Z interpretacji ocen współczynników regresji z tabl. 1 wynika, że wraz ze wzrostem odsetka liczby osób samotnych powyżej 25 roku życia w podregionie rośnie także wartość wskaźnika zagrożenia ubóstwem. Podobną interpretację można także nadać charakterystyce odnoszącej się do udziału gospodarstw domowych z dwiema osobami z wykształceniem co najwyżej zawodowym, jednakże w tym przypadku siła oddziaływania na wartość wskaźnika zagrożenia ubóstwem jest mniejsza. Ocena współczynnika ilorazu liczby osób wymeldowanych do zameldowanych na pobyt stały wskazuje, że w podregionach o wyższej wartości tego wskaźnika obserwuje się także wyższą stopę ubóstwa.

Wymienioną w tabl. 1 zmienną dychotomiczną wprowadzono do modelu w celu wyróżnienia podregionów cechujących się małą gęstością zaludnienia. Jeśli dany podregion znalazł się w grupie podregionów o gęstości zaludnienia mniejszej od 33 percentyla wartości gęstości zaludnienia w podregionach, przyjęto, że wpłynie to negatywnie na wielkość wskaźnika zagrożenia ubóstwem. Kolejne zmienne objaśniające należą do grupy zmiennych opisujących poziom warunków życia gospodarstw domowych. W przypadku obu tych cech — liczby pokoi w przeliczeniu na osobę oraz udziału gospodarstw domowych z wanną lub prysznicem w ogólnej liczbie gospodarstw domowych — przyjmuje się, że wraz ze wzrostem ich wartości poziom ubóstwa w podregionach będzie maleł.

WYNIKI ESTYMACJI

W analizowanym zbiorze danych EU-SILC z 2011 r. w przekroju podregionów liczba zbadanych osób w gospodarstwach domowych była z reguły bardzo mała, zawierała się w przedziale od 174 (podregion ełcki) do 912 (podregion kielecki) osób. Z racji tak małej próby wariancja estymatora bezpośredniego była bardzo duża. Na poziomie podregionów wskaźnik precyzji zawierał się w przedziale 11,2—47,9%. Najwyższy współczynnik zmienności zaobserwowano w podregionie starogardzkim. Niepewność pomiaru jest w tym przypadku bardzo duża, ponieważ przedział 1,0—33,6% pokrywa z prawdopodobieństwem 0,95 prawdziwą wartość wskaźnika zagrożenia ubóstwem. W celu ograniczenia dużej zmienności estymatora bezpośredniego wykorzystano metody statystyki małych obszarów, czyli opisany model Faya-Herriota oraz estymator *EBLUP*.

Przedstawiony w tabl. 1 model został użyty w roli estymatora syntetycznego w estymatorze *EBLUP*. Na podstawie wagi opisanej wzorem (4) został określony udział estymatora bezpośredniego w końcowym szacunku. Po uzyskaniu oszacowań estymatorów oraz błędów standardowych obliczono dodatkowy wskaźnik „zysk na precyzji” wyrażony wzorem:

$$\pi_d = \frac{\sigma(\hat{\theta}_{y,d})}{\sigma(\hat{\theta}_{y,d}^{eblup})} \quad (5)$$

gdzie:

$\sigma(\hat{\theta}_{y,d})$ — błąd standardowy oszacowania estymatora bezpośredniego,

$\sigma(\hat{\theta}_{y,d}^{eblup})$ — błąd standardowy oszacowania estymatora *EBLUP*.

W tabl. 2 przedstawiono oceny estymatorów bezpośredniego oraz *EBLUP* wraz z ich błędami standardowymi, a także wskaźnik „zysk na precyzji”.

TABL. 2. OCENY ESTYMATORA BEZPOŚREDNIEGO ORAZ ESTYMATORA *EBLUP* WSKAŹNIKA ZAGROŻENIA UBÓSTWEM (w %) WRAZ Z BŁĘDAMI STANDARDOWYMI (w p.proc.) ORAZ „ZYSK NA PRECYZJI”

Podregiony	$\hat{\theta}_{y,d}$	$\hat{\theta}_{y,d}^{eblup}$	$\sigma(\hat{\theta}_{y,d})$	$\sigma(\hat{\theta}_{y,d}^{eblup})$	π_d
Jeleniogórski	15,7	17,1	3,4	1,7	2,0
Legnicko-głogowski	14,4	14,5	3,8	1,6	2,3
Wałbrzyski	15,3	20,5	2,9	1,9	1,6
Wrocławski	11,3	12,6	3,2	1,6	2,0
M. Wrocław	6,2	7,5	1,9	1,4	1,3
Bydgosko-toruński	11,5	12,1	2,9	1,5	2,0
Grudziądzki	26,1	22,9	4,4	1,9	2,3
Włocławski	18,3	22,6	3,9	1,9	2,1
Białski	35,2	29,4	6,0	2,2	2,8
Chełmsko-zamojski	34,7	30,2	3,9	2,1	1,9
Lubelski	24,0	18,5	3,9	2,1	1,9
Puławski	35,4	29,5	4,9	2,1	2,3
Gorzowski	31,0	16,4	6,1	2,2	2,8

TABL. 2. OCENY ESTYMATORA BEZPOŚREDNIEGO ORAZ ESTYMATORA *EBLUP* WSKAŹNIKA ZAGROŻENIA UBÓSTWEM (w %) WRAZ Z BŁĘDAMI STANDARDOWYMI (w p.proc.) ORAZ „ZYSK NA PRECYZJI” (dok.)

Podregiony	$\hat{\theta}_{y,d}$	$\hat{\theta}_{y,d}^{ebulp}$	$\sigma(\hat{\theta}_{y,d})$	$\sigma(\hat{\theta}_{y,d}^{ebulp})$	π_d
Zielonogórski	21,7	17,7	4,2	1,8	2,3
Łódzki	14,1	15,1	3,5	1,8	1,9
M. Łódź	13,9	14,2	2,9	1,8	1,6
Piotrkowski	23,6	21,6	3,7	1,8	2,0
Sieradzki	21,5	24,4	4,3	1,8	2,4
Skiernewicki	21,5	23,4	4,5	1,8	2,5
Krakowski	17,7	17,4	3,9	2,0	2,0
M. Kraków	8,4	8,7	2,4	1,5	1,6
Nowosądecki	28,8	23,2	4,7	2,3	2,1
Oświęcimski	12,0	14,3	3,2	1,6	2,0
Tarnowski	40,9	24,6	5,5	2,6	2,1
Ciechanowsko-płocki	18,2	21,3	3,6	1,8	2,1
Ostrołęcko-siedlecki	21,1	25,7	3,4	1,8	1,9
Radomski	23,5	24,5	3,8	2,0	1,9
M. Warszawa	6,2	6,3	1,3	1,1	1,2
Warszawski wschodni	12,8	14,4	2,8	1,7	1,7
Warszawski zachodni	10,8	10,3	2,1	1,4	1,5
Nyski	12,2	16,5	3,4	1,8	1,9
Opolski	14,2	11,5	3,7	1,8	2,0
Krośnieński	25,9	24,1	5,1	1,9	2,7
Przemyski	28,6	26,1	6,1	2,1	3,0
Rzeszowski	14,7	18,0	3,2	1,7	1,9
Tarnobrzeski	19,7	20,9	4,3	1,8	2,4
Białostocki	12,0	13,4	4,2	1,7	2,5
Łomżyński	21,4	24,6	5,2	2,1	2,5
Suwalski	18,5	22,2	7,5	1,9	3,9
Gdański	11,0	11,9	2,9	1,7	1,8
Słupski	29,7	20,8	4,8	2,1	2,3
Starogardzki	17,3	22,0	4,2	1,8	2,3
Trójmiejski	13,3	7,4	3,6	1,9	1,9
Bielski	10,5	11,1	2,2	1,5	1,5
Bytomski	24,1	13,9	5,3	1,9	2,7
Częstochowski	15,2	14,6	3,2	1,6	2,0
Gliwicki	13,4	14,1	3,8	1,7	2,2
Katowicki	13,6	14,6	2,6	1,6	1,7
Rybnicki	10,1	10,4	2,0	1,6	1,3
Sosnowiecki	9,5	10,2	2,1	1,5	1,4
Tyski	10,3	9,9	2,9	1,6	1,9
Kielecki	22,2	21,3	3,5	1,8	2,0
Sandomiersko-jędrzejowski	34,0	29,8	5,9	2,0	2,9
Elbląski	17,6	20,7	4,0	1,8	2,2
Elcki	17,5	20,8	6,3	1,8	3,4
Olsztyński	14,8	17,2	3,4	1,8	1,9
Kaliski	17,5	16,7	3,5	1,7	2,1
Koniński	21,3	19,4	3,8	1,7	2,2
Leszczyński	18,0	17,0	5,1	2,2	2,3
Pilski	21,6	19,8	5,7	1,9	3,0
Poznański	13,4	11,0	3,8	1,9	2,0
M. Poznań	7,7	8,5	3,1	2,0	1,6
Koszaliński	21,9	16,6	4,6	2,0	2,3
Stargardzki	17,3	18,7	8,3	2,1	4,0
M. Szczecin	11,6	9,6	4,1	1,7	2,4
Szczeciński	16,5	12,1	6,6	1,8	3,7

Źródło: opracowanie własne na podstawie EU-SILC 2011 i NSP 2011.

Dzięki wykorzystaniu zaproponowanego modelu Faya-Herriota oraz estymatora *EBLUP* udało się ograniczyć zmienność estymatora bezpośredniego poprawiając w ten sposób jego precyzję. Wartości w kolumnie „zysk na precyzji” (5) zawierają się w przedziale 1,2—4,0, co oznacza, że precyzja w najgorszym przypadku wzrosła o 20%, a w najlepszym o 300%. Zaobserwowany trzykrotny wzrost precyzji miał miejsce w przypadku podregionu stargardzkiego, który wcześniej cechował się najszerszym z analizowanych podregionów przedziałem ufności dla wskaźnika zagrożenia ubóstwem. Obecnie przedział ten jest dużo węższy i prawdziwa wartość wskaźnika zagrożenia ubóstwem z prawdopodobieństwem 0,95 znajduje się w przedziale 14,6—22,8%. Średni „zysk na precyzji” dla 66 podregionów wynosi 2,2.

Analizując następnie wskaźniki precyzji obserwuje się także znaczną poprawę. W tabl. 3 przedstawiono liczebność poszczególnych przedziałów dla współczynnika zmienności.

TABL. 3. WARTOŚCI WSPÓŁCZYNNIKA ZMIENNOŚCI DLA OSZACOWANIA WSKAŹNIKA ZAGROŻENIA UBÓSTWEM W PODREGIONACH POLSKI

Współczynniki zmienności	Oszacowanie estymatora bezpośredniego		Oszacowanie estymatora <i>EBLUP</i>	
	liczebność	w %	liczebność	w %
10% i mniej	0	0,0	27	40,9
11—15%	3	4,5	29	43,9
16—20%	21	31,8	8	12,1
21—25%	20	30,3	1	1,5
26% i więcej	22	33,3	1	1,5

Ź r ó d ł o: opracowanie własne.

EBLUP charakteryzuje się znacznie mniejszą zmiennością w porównaniu do estymatora bezpośredniego. Z sytuacji, w której prawie 64% podregionów charakteryzowało się wskaźnikiem precyzji powyżej 20% zredukowano ten odsetek do zaledwie 3,0%. 56 podregionów, które stanowią 85% wszystkich podregionów, cechowało się współczynnikiem zmienności nieprzekraczającym 15%.

Rezultatem przeprowadzonych prac jest kartogram przedstawiający geograficzne zróżnicowanie ubóstwa w Polsce w podregionach (wykr. 2).

Zobrazowanie otrzymanych wyników na kartogramie pozwala wskazać na silne terytorialne zróżnicowanie wskaźnika zagrożenia ubóstwem. Kształtuje się wyraźny podział na Polskę centralną i wschodnią oraz zachodnią. Najmniejszym wskaźnikiem zagrożenia ubóstwem charakteryzowały się podregiony będące dużymi miastami (wyjątkiem była Łódź — 14,2%). Stolica kraju cechowała się najniższą stopą ubóstwa (6,3%). Następne w kolejności to: podregion trójmiejski (7,4%), Wrocław (7,5%), Poznań (8,5%), Kraków (8,7%) oraz Szczecin (9,6%). Najbardziej zagrożone ubóstwem (wskaźnik powyżej 29%) były osoby w gospodarstwach domowych znajdujących się w podregionach: bialskim (29,4%),

puławskim (29,5%), sandomiersko-jędrzejowskim (29,8%) oraz chełmsko-zamojskim (30,2%).

Podsumowanie

Zastosowanie metod statystyki małych obszarów wykorzystujących źródła danych, które nie są obciążone błędem losowym (rejestry administracyjne) lub jest on bardzo mały (Narodowy Spis Powszechny 2011) pozwoliło na estymację wskaźnika zagrożenia ubóstwem na niepublikowanym dotąd poziomie. Dzięki zastosowaniu estymatora *EBLUP* udało się uzyskać szacunki, które charakteryzowały się dostateczną precyzją, w przeciwieństwie do oszacowań estymatora bezpośredniego. Oprócz tych zalet przedstawione podejście charakteryzuje się jednak także wadami. Model stanowi zawsze pewne uproszczenie rzeczywistości i istnieje ryzyko estymacji wartości wskaźnika zagrożenia ubóstwem, który nie będzie odzwierciedlał rzeczywistej sytuacji w danym obszarze. Dlatego otrzymane szacunki należałoby poddać szczegółowej weryfikacji w kontekście sytuacji społeczno-ekonomicznej.

LITERATURA

- Bedi T., Coudouel A., Simler K. (2007), *More Than a Pretty Picture; Using Poverty Maps to Design Better Policies and Interventions*, World Bank, Washington
- Boonstra H. J., Buelens B. (2011), *Model-based estimation*, „Statistics Netherlands”, Hague
- Dehnel G. (2010), *Statystyka małych obszarów jako narzędzie oceny rozwoju ekonomicznego regionów*, Wydawnictwo Akademii Ekonomicznej w Poznaniu
- Dochody i warunki życia ludności Polski (raport z badania EU-SILC 2011)* (2012), GUS
- Fay R. E., Herriot R. A. (1979), *Estimates of income for small places: An application of James-Stein procedures to census data*, „Journal of the American Statistical Association”, No. 74
- Gołata E. (2004), *Estymacja pośrednia na lokalnym rynku pracy*, Wydawnictwo Akademii Ekonomicznej w Poznaniu
- Haughton J., Khandker S. R. (2009), *Handbook on Poverty and Inequality*, World Bank, Washington
- Ludność. Stan i struktura demograficzno-społeczna* (2013), GUS
- Mapy ubóstwa na poziomie podregionów w Polsce z wykorzystaniem estymacji pośredniej. Prace studialne* (2014), GUS
- Pfeffermann D. (2013), *New Important Developments in Small Area Estimation*, Statistical Science, vol. 28, No. 1

SUMMARY

The article discusses methods for estimating poverty indicators at lower levels than the voivodship spatial aggregation based on different sources of data on households (censuses and administrative records). The use of small area statistics methods, using these sources unaffected by random error, allowed to get results having a corresponding precision.

РЕЗЮМЕ

В статье были обсуждены методы оценки показателей бедности на более низких чем воеводство уровнях пространственного обобщения на основе разных источников данных по домашним хозяйствам (переписи и административные реестры). Применение методов статистики малых домэн, использующих эти источники лишены случайной ошибки, позволило получить результаты характеризующиеся соответственной точностью.