

Estymacja pośrednia wskaźników ubóstwa na poziomie powiatów

Łukasz Wawrowski^a

Streszczenie. Dysponowanie szczegółowymi i precyzyjnymi danymi na temat ubóstwa na niskim poziomie agregacji przestrzennej jest ważne dla prowadzenia skutecznej polityki spójności. W Polsce tego typu informacje są gromadzone w ramach badań gospodarstw domowych, prowadzonych przez Główny Urząd Statystyczny, i udostępniane na poziomie kraju, regionów i wybranych grup społeczno-ekonomicznych. Oszacowania bezpośrednie w domenach, których badanie nie obejmuje, są obciążone dużym błędem szacunku. W sytuacji ograniczonej, w skrajnym przypadku zerowej, liczebności próby estymację umożliwia zastosowanie metod statystyki małych obszarów – estymacji pośredniej. Techniki te wykorzystują cechy silnie skorelowane z badanym zjawiskiem, pochodzące ze spisu powszechnego lub z rejestru administracyjnego.

Celem badania omawianego w artykule jest estymacja dwóch wskaźników: stopy ubóstwa i głębokości ubóstwa na poziomie powiatów, z zastosowaniem empirycznej metody bayesowskiej (EB). Pierwszy wskaźnik informuje o skali zjawiska, a drugi – o jego intensywności, więc są one komplementarnymi miarami ubóstwa. W badaniu wykorzystano dane z Europejskiego Badania Dochodów i Warunków Życia przeprowadzonego w 2011 r. oraz Narodowego Spisu Powszechnego Ludności i Mieszkań 2011. Za pomocą metody EB, bazującej na liniowym modelu mieszanym i symulacjach Monte Carlo, uzyskano informacje o wielkości i intensywności ubóstwa na poziomie powiatów. Oszacowane w ten sposób wskaźniki pozwalają na ocenę zróżnicowania ubóstwa na poziomie lokalnym. Ponadto cechują się większą precyzją i zbliżenością z rejestrami administracyjnymi w porównaniu do rezultatów estymacji bezpośredniej.

Słowa kluczowe: ubóstwo, estymacja pośrednia, empiryczna metoda bayesowska

JEL: C13, C51, I32

Indirect estimation of poverty indicators at poviat level

Abstract. The availability of detailed and precise data on poverty at a low level of spatial aggregation is important when pursuing an effective cohesion policy. In Poland, this type of information is gathered during household surveys conducted by Statistics Poland and is made available at country, region, and selected socio-economic group level. Direct estimates relating to domains not included in a survey are burdened with a serious estimation error. In a situation of a limited (or in extreme cases zero) sample size, an estimation becomes possible through the application of small area estimation methods – indirect estimation. These techniques use variables which are strongly correlated with the researched phenomenon and which come from a census or from an administrative register.

The aim of the study discussed in the article is to estimate two indicators: the rate of poverty and the depth of poverty at a poviat level, with the application of the Empirical Bayes (EB) method. The first indicator provides information on the scale of the phenomenon and the other one on its intensity, and so they constitute complementary measures of poverty. The study used data from the European Union Statistics on Income and Living Conditions of 2011 and the National Census of Population and Housing 2011. Information about the scale and intensity of poverty at the poviat level was obtained through the adaptation of the EB method based on the linear mixed model and Monte Carlo simulations. The indicators estimated this way allow for an assessment of the diversity of poverty at a local level. In addition, they are more precise and consistent with administrative registers in comparison to direct estimation results.

Keywords: poverty, small area estimation, Empirical Bayes method

^a Uniwersytet Ekonomiczny w Poznaniu, Instytut Informatyki i Ekonomii Ilościowej.
ORCID: <https://orcid.org/0000-0002-1201-5344>.

1. Wprowadzenie

Badania empiryczne prowadzone w Polsce, przede wszystkim badania reprezentacyjne Głównego Urzędu Statystycznego (GUS): badanie budżetów gospodarstw domowych (BBGD) oraz Europejskie Badanie Dochodów i Warunków Życia (EU-SILC), dostarczają informacji na temat ubóstwa na bardzo ogólnym poziomie. Estymacja na niższym poziomie agregacji przestrzennej z wykorzystaniem klasycznych metod nie jest możliwa ze względu na duże błędy takich oszacowań.

Dostęp do szczegółowych danych jest podstawą funkcjonowania społeczeństwa informacyjnego. Obserwuje się zapotrzebowanie na dane, które mogłyby zostać wykorzystane m.in. przez władze samorządowe do oceny i planowania polityki społecznej. Na podstawie wyników BBGD i EU-SILC można stwierdzić, że sytuacja dochodowa gospodarstw domowych w Polsce poprawia się z roku na rok. Przeciętny miesięczny dochód rozporządzalny *per capita* w 2009 r. wynosił 1114 zł, a w 2019 r. – 1819 zł, co oznacza wzrost o 63% (GUS, 2020). Stopa ubóstwa szacowana na podstawie EU-SILC w 2018 r. wynosiła 14,8%, podczas gdy 10 lat wcześniej – 16,9%. Obserwuje się także coraz mniejsze nierówności dochodowe. W 2008 r. wartość współczynnika Giniego kształtowała się na poziomie 32,0, a w 2018 r. było to 27,8 (GUS, 2019). Wymienione wskaźniki dotyczą całego kraju; sytuacja dochodowa gospodarstw domowych na poziomie lokalnym może być bardziej zróżnicowana.

Rozszerzenie pokrycia informacyjnego bez konieczności podnoszenia kosztów badania, którą spowodowałoby m.in. zwiększenie wielkości próby, jest możliwe dzięki zastosowaniu metod statystyki małych obszarów – estymacji pośredniej. Umożliwiają one estymację parametrów w przypadku ograniczonej, a nawet zerowej liczebności próby, przy wykorzystaniu wszystkich dostępnych źródeł danych (Żądło, 2015), z tym że wymagają identyfikacji zbiorów danych stanowiących odpowiednie źródło zmiennych pomocniczych w estymacji pośredniej. Cechy te powinny być silnie skorelowane z analizowanym zjawiskiem, a zatem w przypadku analizy ubóstwa poszukiwane są jego ilościowe determinanty (Panek, 2011). Oprócz badań nad ubóstwem (Chambers i Tzavidis, 2006; Molina i Rao, 2010; Wawrowski, 2016) metody statystyki małych obszarów znajdują zastosowanie m.in. w szacowaniu charakterystyk rynku pracy (Gołata, 2004; Wilak, 2014) oraz w statystyce gospodarczej (Chandra, Chambers i Salvati, 2012; Dehnel, 2010).

Badania nad wykorzystaniem metod estymacji pośredniej poziomu ubóstwa są prowadzone także w ramach krajowych i międzynarodowych projektów badawczych. W latach 2013 i 2014 GUS wraz z Bankiem Światowym realizowały projekt *Mapy ubóstwa na poziomie podregionów w Polsce z wykorzystaniem estymacji pośredniej* (Szymkowiak, Młodak i Wawrowski, 2017). Jego bezpośrednim rezultatem było opracowanie raportu *Pomiar ubóstwa na poziomie powiatów (LAU 1)* (Szym-

kowiak, 2015). Estymatorami statystyki małych obszarów posłużono się do oszacowania wartości stopy ubóstwa dla lat 2005, 2008 i 2011. W badaniu omawianym w niniejszym artykule przyjęto szersze podejście – włączono dodatkowy wskaźnik ubóstwa, diagnostykę modelu oraz ocenę wyników. Zastosowano także inną metodę transformacji zmiennej zależnej, z uwzględnieniem aktualnego stanu badań (Rojas-Perilla, Pannier, Schmid i Tzavidis, 2020).

Celem badania omawianego w artykule jest estymacja dwóch wskaźników: stopy ubóstwa i głębokości ubóstwa na poziomie powiatów, z zastosowaniem empirycznej metody bayesowskiej (Molina i Rao, 2010), bazującej na liniowym modelu mieszanym i symulacjach Monte Carlo. Dane wykorzystane w analizie pochodzą z badania EU-SILC przeprowadzonego w 2011 r. oraz Narodowego Powszechnego Spisu Ludności i Mieszkań (NSP) 2011. Otrzymane wyniki poddano wielopłaszczyznowej ocenie z użyciem danych z rejestrów administracyjnych oraz miar precyzji i autokorelacji przestrzennej.

2. Pomiar ubóstwa

Ubóstwo jest uznawane za jeden z najważniejszych i najbardziej złożonych problemów społecznych współczesnego świata. Występuje w każdym społeczeństwie. Kluczowy w ocenie tego zjawiska jest jego pomiar. Nadanie mu uniwersalnego charakteru wymagałoby określenia jednolitych zasad i definicji. Rada Europejskiej Wspólnoty Gospodarczej (WE) w dokumencie z 19 grudnia 1984 r. zdefiniowała je następująco: „ubóstwo odnosi się do osób, rodzin lub grup osób, których zasoby (materialne, kulturowe i społeczne) są ograniczone w takim stopniu, że poziom ich życia obniża się poza akceptowalne minimum w kraju zamieszkania” (EEC, 1985).

W powyższej definicji kluczowym aspektem jest ustalenie kryterium przynależności do sfery ubóstwa. Można zastosować kryterium jedno- lub wielowymiarowe. Za pomocą tego pierwszego ocenia się stopień zaspokojenia potrzeb jednostki w odniesieniu do dochodów (wydatków) wyrażonych w formie monetarnej. W literaturze przedmiotu podkreśla się jednak niedoskonałość tego podejścia i wyraża się przekonanie, że zjawisko ubóstwa nie jest jednowymiarowe (GUS i US w Łodzi, 2013a; Panek, 2010). Wielu badaczy postulowało uwzględnienie w jego analizie także czynników pozamonetarnych, m.in. dochodów i zasobów materialnych nagromadzonych we wcześniejszych okresach, warunków mieszkaniowych, edukacji oraz zasobów zawodowych i finansowych (Panek, 2011; Ulman i Ćwiek, 2020).

Podejście jednowymiarowe wymaga określenia poziomu dochodów (wydatków), poniżej którego osoba jest uznawana za ubogą. W Polsce zdefiniowano kilka rodzajów granicy (linii) ubóstwa. Instytut Pracy i Spraw Socjalnych (IPiSS) wyznacza wartości minimum egzystencji (granicy wydatków pozwalających na skromne wyżywienie oraz utrzymanie małego mieszkania) i minimum socjalnego (granicy wydatków

umożliwiająca minimalnie godziwy standard życia), a także bierze udział w konsultacji ustawowej granicy ubóstwa (kwoty dochodu, która uprawnia do ubiegania się o przyznanie świadczenia z systemu pomocy społecznej) (Kurowski, 2002). W badaniach reprezentacyjnych granicę ubóstwa ustala się jako stałą część mediany lub średniej arytmetycznej rozkładu dochodów w całej populacji. Główny Urząd Statystyczny w BBGD jako linię ubóstwa przyjmuje 50% średnich wydatków. Natomiast w EU-SILC określono, za rekomendacją Eurostatu, że granicę ubóstwa stanowi 60% mediany rozkładu dochodów ekwiwalentnych (uwzględniających skład demograficzny gospodarstwa domowego). Tę definicję przyjęto także w niniejszej pracy.

2.1. Wskaźniki ubóstwa

Pomiar ubóstwa bazuje na dwóch podstawowych wskaźnikach: stopie ubóstwa i głębokości ubóstwa. Pierwsza miara, informująca o odsetku ubogich w populacji, jest prosta w budowie i interpretacji, ale niepozbawiona wad. Przede wszystkim nie uwzględnia intensywności ubóstwa. Osoby znajdujące się poniżej granicy ubóstwa mogą mieć dochody bardzo zbliżone do tej granicy lub bardzo od niej oddalone, ale w obu przypadkach stopa ubóstwa będzie taka sama, mimo że zjawisko jest bardziej nasilone w drugiej sytuacji. Intensywność tę odzwierciedla głębokość ubóstwa, która informuje o poziomie ubóstwa wśród osób znajdujących się poniżej linii ubóstwa. Oparcie analizy na obu wymienionych wskaźnikach daje więc pełniejszy obraz zjawiska (Haughton i Khandker, 2009).

Na podstawie wartości dochodów ekwiwalentnych można ustalić odpowiednie miary ubóstwa. Ogólny wzór na wskaźniki ubóstwa wyznaczyli Foster, Greer i Thorbecke (1984); miary te, od inicjałów nazwisk autorów, nazwano FGT. Zakłada się, że dana jest skończona populacja $U = 1, \dots, j, \dots, N$ podzielona na D domen lub obszarów o liczebnościach N_1, \dots, N_D . Wartość granicy ubóstwa określona jest przez z , a E_{dj} oznacza dochód j -tej jednostki w d -tej domenie. Jeśli $E_{dj} < z$, wówczas jednostka j z obszaru d określana jest jako znajdująca się w sferze ubóstwa. Ogólny wzór na wskaźniki ubóstwa z rodziny FGT wyrażony jest następująco:

$$F_{\alpha d} = \frac{1}{N} \sum_{j=1}^{N_d} \left(\frac{z - E_{dj}}{z} \right)^{\alpha} I(E_{dj} < z), \quad \alpha \geq 0, \quad d = 1, \dots, D \quad (1)$$

gdzie $I(E_{dj} < z) = 1$, jeśli $E_{dj} < z$, oraz $I(E_{dj} < z) = 0$ w przeciwnym przypadku.

Dla $\alpha = 0$ otrzymuje się wskaźnik zasięgu ubóstwa, czyli stopę ubóstwa. Z kolei po podstawieniu $\alpha = 1$ wynikiem jest głębokość ubóstwa, która informuje o odległości dochodów osób ubogich od granicy ubóstwa. Innymi słowy, wskaźnik ten określa stopień ubóstwa osób ubogich.

2.2. Determinanty ubóstwa

Działania prowadzące do redukcji poziomu ubóstwa wymagają identyfikacji jego symptomów. Haughton i Khandker (2009) wskazują trzy poziomy, na których można obserwować kluczowe przyczyny ubóstwa i mierzyć charakterystyki tego zjawiska:

- regionalny;
- środowiska lokalnego;
- jednostki (gospodarstwa domowego lub osoby), obejmujący charakterystyki demograficzne, ekonomiczne i społeczne.

Główny Urząd Statystyczny w swoich publikacjach rozpatruje zjawisko ubóstwa w bardzo zbliżonych przekrojach społeczno-demograficzno-ekonomicznych, przy czym wśród typów gospodarstwa domowego wyróżnia także małżeństwa z co najmniej czworgiem dzieci na utrzymaniu, w przypadku których zasięg ubóstwa jest bardzo wysoki i w 2011 r. wynosił 47,5% (GUS i US w Łodzi, 2013b).

Uwzględnienie grup wieku pozwala zauważyć, że wartości stopy ubóstwa w grupie 0–17 lat są wyższe (23,1%) niż wśród osób w wieku 18–64 lat (15,9%) czy 65 lat i więcej (11,2%).

W badaniu *Jakość życia, kapitał społeczny, ubóstwo i wykluczenie społeczne w Polsce* (GUS i US w Łodzi, 2013a) opracowano model regresji logistycznej dla ubóstwa dochodowego. W kategorii głównego źródła utrzymania gospodarstwa domowego największe prawdopodobieństwo znalezienia się w sferze ubóstwa miały gospodarstwa utrzymujące się ze świadczeń społecznych i z rent – w odniesieniu do poziomu referencyjnego, jakim była praca najemna. W omawianym badaniu uwzględniono także zawód głowy gospodarstwa domowego, a za poziom referencyjny przyjęto wartości wskaźników dla robotników przemysłowych i rzemieślników. Stwierdzono, że większe ryzyko ubóstwa dotyczy osób, które pracują jako rolnicy, ogrodnicy, leśnicy i rybacy, a także pracownicy zatrudnieni przy pracach prostych. Najmniejsze prawdopodobieństwo ubóstwa odnotowano wśród kadry menedżerskiej, wyższych urzędników i kierowników.

3. Metody estymacji poziomu ubóstwa

Podstawowym narzędziem w szacowaniu parametrów populacji na podstawie danych z badań reprezentacyjnych jest estymacja bezpośrednia. Przy odpowiednio licznej próbie estymator bezpośredni jest nieobciążony i efektywny (Rao i Molina, 2015), ale na niższym poziomie agregacji, co oznacza mniej liczną próbę, traci swoje własności. Pomimo to oceny parametrów uzyskane za pomocą estymatora bezpośredniego są wykorzystywane jako wielkości wejściowe przy estymacji na poziomie obszaru i stanowią punkt odniesienia w ocenie jakości oszacowań.

3.1. Estymacja bezpośrednia

W badaniach reprezentacyjnych stosuje się estymator bezpośredni zaproponowany przez Horvitz i Thompsona (1952). Niech s oznacza próbę wylosowaną z populacji U , a s_d będzie podpróbą z obszaru d o liczebności $n_d < N_d$. Przez $w_{dj} = \pi_{dj}^{-1}$ oznaczono wagę z próby będącą odwrotnością prawdopodobieństwa inkluzji pierwszego rzędu. Oszacowanie wartości globalnej w domenie d uzyskuje się z wykorzystaniem wzoru:

$$\hat{Y}_d^{HT} = \sum_{j=1}^{n_d} y_{dj} w_{dj} \quad (2)$$

gdzie:

\hat{Y}_d^{HT} – oszacowanie globalnej wartości cechy Y w d -tej domenie,

y_{dj} – wartość cechy Y dla j -tej jednostki w d -tym obszarze,

w_{dj} – wartość wagi wynikającej ze schematu losowania dla j -tej jednostki w d -tym obszarze.

W celu estymacji bezpośredniej wskaźników z rodziny FGT należy zmodyfikować formułę (2) w następujący sposób:

$$\hat{F}_{\alpha d}^{HT} = N_d^{-1} \sum_{j \in s_d} w_{dj} \left(\frac{z - E_{dj}}{z} \right)^{\alpha} I(E_{dj} < z), \quad j = 1, \dots, N_d \quad (3)$$

gdzie:

w_{dj} – wartość wagi wynikającej ze schematu losowania dla j -tej jednostki w d -tym obszarze,

N_d – liczba osób w d -tym obszarze,

E_{dj} – dochód j -tej jednostki,

z – granica ubóstwa,

$I(\cdot)$ – funkcja indykatorowa przyjmująca wartość 1, jeśli wyrażenie wewnątrz funkcji jest prawdziwe, a wartość 0 w przeciwnym przypadku.

Estymator bezpośredni wykorzystuje wyłącznie dane pochodzące z próby i dla dużych wartości n_d jest nieobciążony i efektywny, natomiast mała liczebność próby implikuje duże wartości wariancji. Jeśli dana domena w próbie nie jest w ogóle reprezentowana ($n_d = 0$), to ten sposób estymacji nie ma zastosowania (Guadarrama, Molina i Rao, 2016).

3.2. Estymacja pośrednia

W metodach estymacji pośredniej przyjmuje się podejście obszarowe lub jednostkowe. W podejściu obszarowym zmienną zależną jest wskaźnik ubóstwa mierzony na określonym poziomie terytorialnym, a w podejściu jednostkowym – dochód gospodarstwa domowego. W omawianym badaniu zastosowano ten drugi sposób.

3.2.1. Empiryczna metoda bayesowska

Empiryczna metoda bayesowska (EB) polega na dopasowaniu liniowego modelu mieszanego opisującego dochód na poziomie gospodarstwa domowego na podstawie danych z badania reprezentacyjnego, a następnie na przeprowadzeniu dużej liczby symulacji Monte Carlo z wykorzystaniem danych z badania pełnego w celu wyznaczenia teoretycznych wartości dochodu dla osób spoza próby na podstawie zmiennej pomocniczych (Molina i Rao, 2010). Metoda ta jest wykorzystywana do estymacji ubóstwa m.in. przez Bank Światowy (Van der Weide, 2014).

Dany jest wektor $y = (y_1, \dots, y_N)'$, zawierający wartości zmiennej losowej powiązanej z N jednostkami skończonej populacji. Niech y_s będzie fragmentem wektora y zawierającym informacje z wylosowanej próby s , a y_r – wektorem elementów spoza próby. Po uporządkowaniu elementów można zapisać, że $y = (y'_s, y'_r)$. Celem jest oszacowanie wartości funkcji $\delta = h(y)$ zmiennej losowej y z wykorzystaniem danych z próby y_s . Dla estymatora $\hat{\delta}$ średni kwadrat błędu (MSE) jest dany wzorem:

$$\text{MSE}(\hat{\delta}) = E_y\{(\hat{\delta} - \delta)^2\} \quad (4)$$

gdzie E_y – wartość oczekiwana z uwzględnieniem łącznego rozkładu wektora y .

Najlepszym estymatorem (ang. *best predictor*, BP) parametru δ jest funkcja y_s , która minimalizuje (4) i jest dana warunkową wartością oczekiwaną:

$$\hat{\delta}^B = E_{y_r}(\delta | y_s) \quad (5)$$

gdzie wartość oczekiwana uwzględnia warunkowy rozkład y_r .

Warto zauważyć, że ten estymator jest nieobciążony, ponieważ:

$$E_{y_s}(\hat{\delta}^B) = E_{y_s}\{E_{y_r}(\delta | y_s)\} = E_y(\delta) \quad (6)$$

Zwykle $\hat{\delta}^B$ zależy od wektora θ nieznanymi parametrów modelu. Wówczas empiryczny BP $\hat{\delta}$ otrzymuje się poprzez zastąpienie θ odpowiednim estymatorem $\hat{\theta}$ i następnie oszacowanie (5), przyjmując $\theta = \hat{\theta}$.

Poniżej zaprezentowano sposób oszacowania BP dla rodziny wskaźników ubóstwa FGT w przypadku małych domen. Zakłada się transformację zmiennej dochodowej E_{dj} według formuły $Y_{dj} = T(E_{dj})$, biorąc pod uwagę, że wektor y zawiera wartości transformowanej zmiennej Y_{dj} dla wszystkich jednostek populacji, tak że $y \sim N(\mu, V)$. Wówczas zmienną losową

$$\hat{F}_{adj} = \left(\frac{z - E_{dj}}{z} \right)^\alpha I(E_{dj} < z), \quad j = 1, \dots, N_d \quad (7)$$

można wyrazić w kontekście Y_{dj} jako:

$$F_{adj} = \left(\frac{z - T^{-1}(Y_{dj})}{z} \right)^\alpha I\{T^{-1}(Y_{dj}) < z\} =: h_\alpha(Y_{dj}), \quad j = 1, \dots, N_d \quad (8)$$

Wobec tego miara ubóstwa (1) jest nieliniową funkcją wektora y . Po przyjęciu za $\delta = F_{adj}$ i podstawieniu do (5) najlepszy estymator (BP) parametru F_{adj} można wyrazić wzorem:

$$\hat{F}_{adj}^B = E_{y_r}(F_{adj}|y_s) \quad (9)$$

Dekomponując F_{adj} (1) na jednostki wylosowane i spoza próby oraz uwzględniając warunkową wartość oczekiwaną (9), otrzymuje się estymator wskaźnika ubóstwa:

$$\hat{F}_{adj}^B = \frac{1}{N_d} \left\{ \sum_{j \in s_d} F_{adj} + \sum_{j \in r_d} \hat{F}_{adj}^B \right\} \quad (10)$$

gdzie \hat{F}_{adj}^B jest najlepszym estymatorem $F_{adj} = h_\alpha(Y_{dj})$, danym przez

$$\hat{F}_{adj}^B = E_{y_r}[h_\alpha(Y_{dj})|y_s] = \int_{IR} h_\alpha(y) f_{Y_{dj}}(y|y_s) dy, \quad j \in r_d \quad (11)$$

gdzie $f_{Y_{dj}}(y \vee y_s)$ jest warunkową gęstością Y_{dj} wektora y_s .

Wartość oczekiwana (11) nie może zostać wyliczona bezpośrednio z powodu złożoności $h_\alpha(y)$. Jednakże w przypadku gdy $y = (y'_s, y'_r)'$ ma rozkład normalny ze średnią daną wektorem $\mu = (\mu'_s, \mu'_r)'$, a rozkład y_r pod warunkiem y_s dany jest przez

$$y_r|y_s \sim N(\mu_{r|s}, V_{r|s}) \quad (12)$$

gdzie:

$$\mu_{r|s} = \mu_r + V_{rs}V_s^{-1}(y_s - \mu_s) \quad \text{i} \quad V_{r|s} = V_r - V_{rs}V_s^{-1}V_{sr} \quad (13)$$

to empiryczne przybliżenie rozwiązania (11) jest możliwe z wykorzystaniem symulacji Monte Carlo o dużej liczbie L wektorów y_r wygenerowanych z (12).

Niech $Y_{dj}^{(l)}$ oznacza wartości zmiennej Y_{dj} , $j \in r_d$ spoza próby, otrzymane w l symulacjach $l = 1, \dots, L$. Przybliżenie Monte Carlo najlepszego estymatora (BP) Y_{dj} dla $j \in r_d$ można wyrazić jako:

$$\hat{F}_{\alpha dj}^B = E_{y_r}[h_\alpha(Y_{dj})|y_s] \approx \frac{1}{L} \sum_{\ell=1}^L h_\alpha(Y_{dj}^{(\ell)}), \quad j \in r_d \quad (14)$$

Końcowy estymator to $\hat{F}_{\alpha dj}^{EB}$, będący najlepszym empirycznym estymatorem (EBP) parametru $F_{\alpha dj}$. Ostatecznie EBP miary ubóstwa $F_{\alpha d}$ jest określony wzorem:

$$\hat{F}_{\alpha d}^{EB} = \frac{1}{N_d} \left\{ \sum_{j \in s_d} F_{\alpha dj} + \sum_{j \in r_d} \hat{F}_{\alpha dj}^{EB} \right\} \quad (15)$$

Wartość oczekiwana (9) może być bezpośrednio przybliżona przez symulację Monte Carlo, co pozwala na estymację właściwie dowolnego parametru dla domeny $\delta_d = h(y_d)$. Wówczas procedura estymacji parametru dla domeny $\delta_d = h(y_d)$ za pomocą metody EB jest następująca:

1. estymacja nieznanymi parametrów rozkładu θ transformowanego wektora y z wykorzystaniem danych z próby y_s ;
2. wygenerowanie L wektorów spoza próby $y_r^{(l)}$, $l = 1, \dots, L$ na podstawie (12) i (13); θ jest zastępowane estymatorem $\hat{\theta}$ uzyskanym w pkt 1;
3. rozszerzenie każdego z L wygenerowanych wektorów $y_r^{(l)}$ danymi z próby y_s do postaci wektora populacji $y^l = (y'_s, (y_r^{(l)})')'$, $l = 1, \dots, L$;
4. wyznaczenie wartości parametru dla domeny $\delta_d^{(l)} = h(y_d^{(l)})$ z wykorzystaniem elementów wektora $y^{(l)}$ dla d -tej domeny - $y_d^l = (y'_{ds}, (y'_{dr})')$;
5. przybliżenie Monte Carlo estymatora EBP parametru δ_d poprzez uśrednienie wartości parametru dla domeny na podstawie L pseudopopulacji:

$$\delta_d^{EB} = \frac{1}{L} \sum_{\ell=1}^L \delta_d^{(\ell)} \quad (16)$$

3.2.2. Liniowy model regresji z zagnieżdżonym składnikiem losowym

Model nadpopulacji ξ oparty na liniowym modelu mieszanym (Battese, Harter i Fuller, 1988) może zostać wykorzystany do oszacowania (15). Model opisuje liniową dla wszystkich domen relację pomiędzy transformowaną zmienną dochodową Y_{dj} a wektorem x_{dj} , zawierającym wartości p zmiennych niezależnych, a także uwzględnia efekt losowy dla domeny u_d wraz z resztami e_{dj} :

$$\xi: Y_{dj} = x'_{dj}\beta + u_d + e_{dj}, \quad j = 1, \dots, N_d, \quad d = 1, \dots, D \quad (17)$$

gdzie $u_d \stackrel{iid}{\sim} N(0, \sigma_u^2)$ i $e_{dj} \stackrel{iid}{\sim} N(0, \sigma_e^2)$, a u_d oraz e_{dj} są niezależne.

Wektory i macierze otrzymuje się poprzez wydzielenie elementów dla domeny d :

$$y_d = \text{col}_{1 \leq j \leq N_d}(Y_{dj}), \quad e_d = \text{col}_{1 \leq j \leq N_d}(e_{dj}), \quad X_d = \text{col}_{1 \leq j \leq N_d}(x'_{dj}) \quad (18)$$

Wówczas wektor $y_d, d = 1, \dots, D$ jest niezależny z $y_d \sim N(\mu_d, V_d)$, gdzie:

$$\mu_d = X_d\beta \quad \text{ i } \quad V_d = \sigma_u^2 \mathbf{1}_{N_d} \mathbf{1}'_{N_d} + \sigma_e^2 I_{N_d} \quad (19)$$

przy czym $\mathbf{1}_k$ oznacza kolumnowy wektor jedynek o rozmiarze k , a I_k jest macierzą jednostkową $k \times k$.

Rozważana jest dekompozycja y_d na jednostki wylosowane i te spoza próby $y_d = (y'_{ds}, y'_{dr})'$ w przypadku $n_d > 0$ i podobnie dla X_d, μ_d i V_d . Przy założeniu normalności modelu (17) można zauważyć, że $y_{dr} \vee y_{ds} \sim N(\mu_{drvs}, V_{drvs})$, gdzie:

$$\mu_{dr|s} = X_{dr}\beta + \sigma_u^2 \mathbf{1}_{N_d - n_d} \mathbf{1}'_{n_d} V_{ds}^{-1} (y_{ds} - X_{ds}\beta) \quad (20)$$

i

$$V_{dr|s} = \sigma_u^2 (1 - \gamma_d) \mathbf{1}_{N_d - n_d} \mathbf{1}'_{N_d - n_d} + \sigma_e^2 I_{N_d - n_d} \quad (21)$$

dla $V_{ds} = \sigma_u^2 \mathbf{1}_{n_d} \mathbf{1}'_{n_d} + \sigma_e^2 I_{n_d}$ oraz $\gamma_d = \sigma_u^2$.

Zastosowanie przybliżenia Monte Carlo (14) pociąga za sobą symulację D wektorów y_{dr} o rozmiarze $N_d - n_d, d = 1, \dots, D$ i wielowymiarowym rozkładzie normalnym. Powtórzenie tego procesu L razy sprawia, że jest on wymagający obliczeniowo, a dla dużych N_d - niewykonalny. Można tego uniknąć, gdy się zauważy, że macierz V_{drvs} (21) odpowiada macierzy kowariancji wektora y_{dr} wygenerowanego przez model

$$y_{dr} = \mu_{dr|s} + v_d \mathbf{1}_{N_d - n_d} + \varepsilon_{dr} \quad (22)$$

z nowym efektem losowym v_d oraz błędem ε_{dr} , które są niezależne, o rozkładach:

$$v_d \sim N\{0, \sigma_u^2(1 - \gamma_d)\}, \quad d = 1, \dots, D, \quad \varepsilon \sim N(0_{N_d - n_d}, \sigma_e^2 I_{N_d - n_d}) \quad (23)$$

Jeśli wykorzystuje się (22), to zamiast generowania wektora y_{dr} o wielowymiarowym rozkładzie normalnym wystarczy wygenerować jednowymiarowe zmienne $v_d \sim N\{0, \sigma_u^2(1 - \gamma_d)\}$ i $\varepsilon_{dj} \sim N(0, \sigma_e^2)$ niezależnie dla $j \in r_d$. Wówczas uzyskuje się odpowiadające elementy wektora $y_{dj}, j \in r_d$ z (22) przy wykorzystaniu μ_{drvs} danego przez (23). Jak już wspomniano, w praktyce parametry modelu $\theta = (\beta', \sigma_u^2, \sigma_e^2)'$ są zastępowane przez odpowiednie estymatory $\hat{\theta} = (\hat{\beta}', \hat{\sigma}_u^2, \hat{\sigma}_e^2)'\hat{\theta}'$ i wówczas wartości zmiennej Y_{dj} są generowane z rozkładu normalnego o odpowiednim odchyleniu standardowym.

Jeśli domena d nie znalazła się w próbie, to wartości $Y^{(\ell)aj}$ dla $j = 1, \dots, N_d$ są generowane metodą bootstrap z wykorzystaniem $Y_{dj} = x'_{dj}\hat{\beta} + u_d^* + e_{dj}^*$, gdzie $u_d^* \stackrel{iid}{\sim} N(0, \hat{\sigma}_u^2)$ i $e_{dj}^* \stackrel{iid}{\sim} N(0, \hat{\sigma}_e^2)$ oraz u_d^* jest niezależne od e_{dj}^* . Wzór (14) wykorzystuje się wówczas do otrzymania estymatora \hat{F}_{adj}^{EB} dla F_{adj} i estymatora EB F_{ad} jako:

$$\hat{F}_{ad}^{EB} = N_d^{-1} \sum_{j=1}^{N_d} \hat{F}_{adj}^{EB} \quad (24)$$

Estymator (24) ma charakter syntetyczny, jeśli żadna jednostka nie jest obserwowana w domenie d .

Opisane podejście bazuje na transformacji zmiennej zależnej w celu zbliżenia rozkładu cechy do rozkładu normalnego. W przypadku występowania silnej skośności operacja ta może nie przynieść pożądanych efektów. Można wtedy zastosować model wykorzystujący uogólniony rozkład beta (Graf, Marín i Molina, 2019), ale wiąże się to ze skomplikowaniem obliczeń.

3.3. Ocena jakości oszacowań

Jako miarę precyzji oszacowań uzyskanych za pomocą opisanych metod przyjęto pierwiastek względnego średniego kwadratu błędu oszacowania – RRMSE (ang. *relative root mean square error*), dany wzorem:

$$\text{RRMSE}(\hat{F}_{ad}) = \frac{\sqrt{\text{MSE}(\hat{F}_{ad})}}{\hat{F}_{ad}} \quad (25)$$

gdzie:

$\sqrt{\text{MSE}(\hat{F}_{ad})}$ – średni błąd oszacowania wskaźnika ubóstwa,
 \hat{F}_{ad} – oszacowanie wskaźnika ubóstwa.

Wartości średniego błędu oszacowania można aproksymować z wykorzystaniem metody bootstrap. W przypadku estymacji bezpośredniej przegląd podejść można znaleźć w pracy Woltera (2007), natomiast w przypadku metody EB odpowiedni algorytm zaproponowali González-Manteiga, Lombardia, Molina, Morales i Santamaría (2008).

4. Estymacja pośrednia ubóstwa na poziomie powiatów

W EU-SILC zrealizowanym w 2011 r. wzięło udział 12871 gospodarstw domowych z 375 spośród wszystkich 379 powiatów (cztery powiaty nie były reprezentowane w próbie). Analiza prowadzona na poziomie powiatów wykazała, że w kolejnych 12 powiatach nie znalazła się ani jedna osoba uboga, co uniemożliwiło estymację bezpośrednią dla tych jednostek terytorialnych.

Kolejnym etapem badania było oszacowanie wskaźników ubóstwa na poziomie powiatów. Z uwagi na duże błędy standardowe bezpośrednich oszacowań wskaźników ubóstwa, a tym samym ich małą precyzję, zrezygnowano z podejścia klasycznego i zastosowano estymację pośrednią. Estymacja charakterystyk ubóstwa z wykorzystaniem metod statystyki małych obszarów wymaga – w celu poprawy jakości – użycia zmiennych pomocniczych. Cechy te powinny pochodzić ze źródeł pozbawionych błędów losowych, np. z badań pełnych bądź rejestrów administracyjnych. W podejściu modelowym kluczowy jest dobór zmiennych, które dobrze opisują zróżnicowanie estymowanych cech. Nieodpowiednia specyfikacja zmiennych niezależnych może powodować obciążenie wyników.

4.1. Liniowy model mieszany dochodu ekwiwalentnego

Na podstawie danych z EU-SILC z 2011 r. oszacowano parametry liniowego modelu regresji z zagnieżdżonym składnikiem losowym. Wartości zmiennej zależnej – dochodu ekwiwalentnego gospodarstwa – zostały przekształcone za pomocą transformacji Boxa-Coxa. Jako zmienne pomocnicze wykorzystano cechy demograficzne, ekonomiczne i społeczne mierzone na poziomie gospodarstwa domowego, które wybrano na podstawie studiów literatury dotyczących symptomów ubóstwa. Za efekt losowy przyjęto powiat. Obliczenia wykonano w programie R z wykorzystaniem pakietu lme4 (Bates, Mächler, Bolker i Walker, 2015). Parametry modelu przedstawia tabl. 1¹.

¹ Dane zaprezentowane w tablicy są tożsame z wartościami zamieszczonymi w raporcie *Pomiar ubóstwa na poziomie powiatów (LAU 1)* (Szymkowiak, 2015) ze względu na zastosowanie innej transformacji zmiennej zależnej.

Tabl. 1. Parametry liniowego modelu mieszanego objaśniającego dochód ekwiwalentny na poziomie gospodarstwa domowego

Zmienne niezależne	Parametr β	Błąd standardowy	Statystyka t
(wyraz wolny)	27,65	0,09	319,37
Odsetek mężczyzn w gospodarstwie	0,79	0,10	7,76
Odsetek osób w wieku: 30–44 lat w gospodarstwie	0,56	0,12	4,58
65 lat i więcej w gospodarstwie	-0,32	0,08	-3,91
Odsetek osób: bezrobotnych w gospodarstwie	-5,05	0,19	-26,54
niepełnosprawnych w gospodarstwie	-2,44	0,16	-15,15
z wykształceniem podstawowym w gospodarstwie	-1,00	0,10	-10,30
z wykształceniem wyższym w gospodarstwie	4,05	0,11	38,50
Gospodarstwo posiadające: 1 pokój (zmienna binarna)	-0,89	0,10	-9,26
3 pokoje i więcej (zmienna binarna)	0,77	0,06	13,10
Miejsce zamieszkania: wieś lub miasto do 20 tys. mieszkańców (zmienna binarna)	-1,11	0,06	-17,11
Wskaźnik obciążenia demograficznego dzieci w gospodarstwie	-1,07	0,07	-14,69

Źródło: opracowanie własne na podstawie danych z EU-SILC 2011.

Wszystkie zmienne niezależne są istotne, a odpowiadające im znaki przy parametrze β są zgodne z logicznym kierunkiem wpływu na zmienną zależną. Większe wartości czterech zmiennych: odsetka mężczyzn w gospodarstwie, odsetka osób w wieku 30–44 lat w gospodarstwie, odsetka osób z wykształceniem wyższym w gospodarstwie oraz gospodarstwo posiadające 3 pokoje i więcej (zmienna binarna) wpływają na wzrost dochodu w gospodarstwie. Wartości parametru β przy pozostałych zmiennych mają znak ujemny, co oznacza, że większe wartości tych cech powodują spadek dochodu w gospodarstwie.

Wariancja efektu losowego wyniosła $\widehat{\sigma}_u = 0,2702$, natomiast wariancja błędu losowego $\widehat{\sigma}_e = 8,7083$. Na podstawie testu permutacyjnego wykazano, że efekt losowy jest w modelu istotny. Przeprowadzono diagnostykę modelu z wykorzystaniem wykresów kwantylowych. Stwierdzono, że zarówno rozkład efektów, jak i reszt jest zbliżony do normalnego. Pozytywnie zweryfikowano także założenie o niezależności efektów losowych i reszt (Biecek, 2011).

W kolejnym kroku zastosowano algorytm estymacji metodą EB dla danych pochodzących z NSP 2011; przyjęto $L = 200$ i $B = 200$. Obliczenia przeprowadzono z wykorzystaniem pakietu sae (Molina i Marhuenda, 2015) w programie R. Na tej podstawie uzyskano oszacowania stopy i głębokości ubóstwa na poziomie powiatów w Polsce. Tablica 2 zawiera rozkład względnych błędów oszacowań stopy i głębokości ubóstwa.

Tabl. 2. Rozkład względnych błędów oszacowań stopy i głębokości ubóstwa oszacowanych z wykorzystaniem estymacji bezpośredniej (HT) i pośredniej (EB)

Wskaźniki i metody estymacji	Minimum	Q1	Mediana	Średnia	Q3	Maksimum
	w %					
Stopa ubóstwa: HT	11,50	22,28	28,52	32,81	37,57	99,97
EB	6,45	10,36	11,58	11,81	13,08	19,31
Głębokość ubóstwa: HT	14,41	25,05	31,90	37,00	42,46	99,97
EB	9,77	13,19	14,46	14,78	16,17	23,62

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

Estymacja pośrednia poprawia precyzję oszacowań w porównaniu do estymacji bezpośredniej. Średni wskaźnik precyzji dla stopy ubóstwa oszacowanej z wykorzystaniem metody EB wyniósł 11,81%, podczas gdy dla estymacji bezpośredniej było to 32,81%. Podobną poprawę średniej wartości RRMSE można zaobserwować także w przypadku głębokości ubóstwa – z 37,00% (estymacja HT) do 14,78% (estymacja EB). Względny błąd oszacowania stopy ubóstwa nie przekroczył 20%.

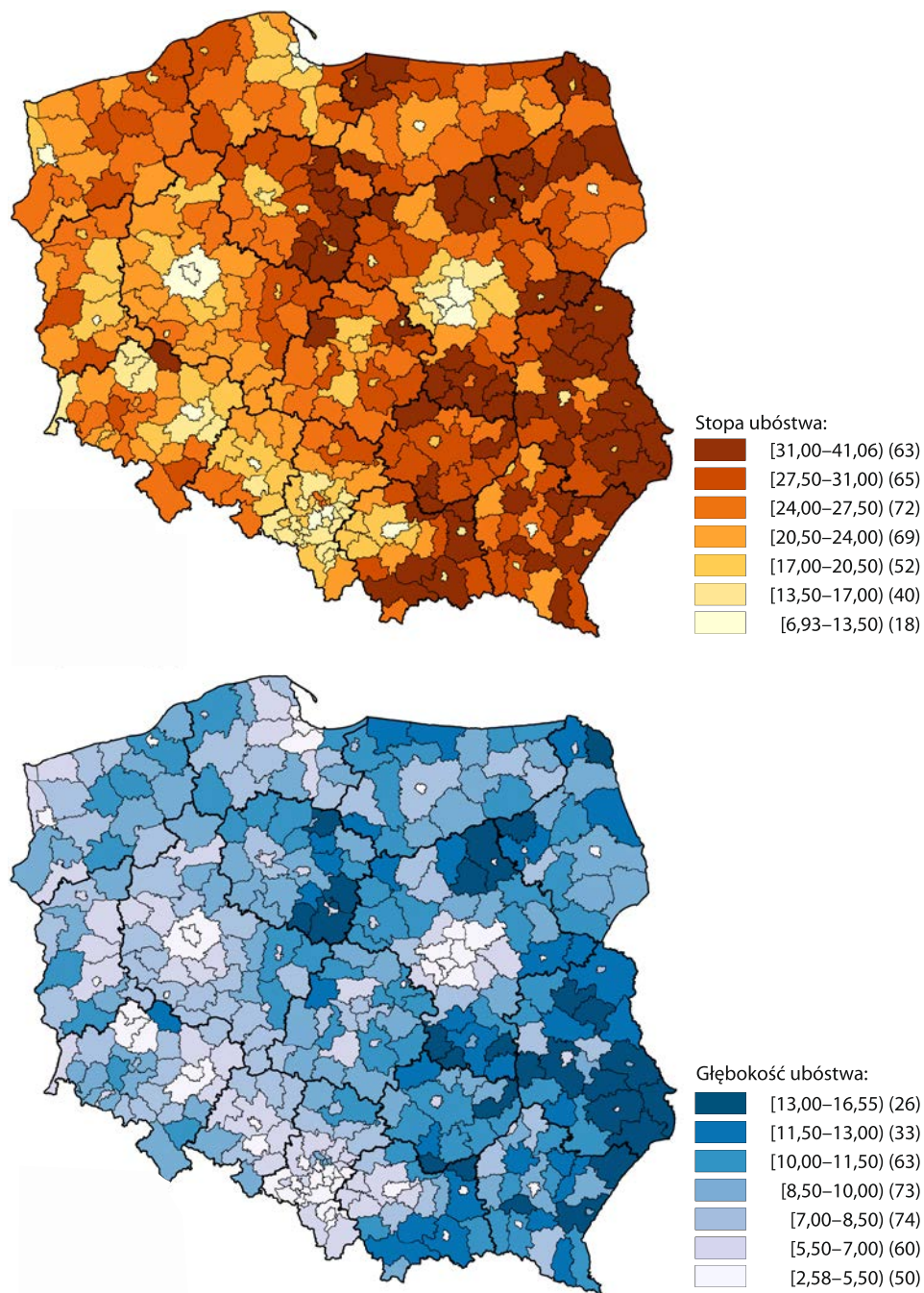
4.2. Terytorialne zróżnicowanie ubóstwa w Polsce

Przeprowadzono także przestrzenną analizę ubóstwa w Polsce. Jej rezultaty przedstawiono na kartogramach (mapa), które pozwalają na zobrazowanie przestrzennych zależności (Bedi, Coudouel i Simler, 2007). Zidentyfikowano obszary najmniej i najbardziej narażone na występowanie zjawiska ubóstwa. Ponadto przeprowadzono ocenę spójności przestrzennej otrzymanych oszacowań wskaźników ubóstwa z wykorzystaniem statystyki Morana I.

Największe wartości stopy ubóstwa obserwuje się w powiatach we wschodniej części kraju oraz jednostkach położonych w znacznej odległości od stolic województw. Występowaniem największego ubóstwa – powyżej 40% – charakteryzują się powiaty chełmski (41,1%) i hrubieszowski (40,9%). Wśród 20 powiatów o największych wartościach stopy ubóstwa znajdują się trzy powiaty położone we wschodniej części woj. kujawsko-pomorskiego, pięć powiatów woj. lubelskiego i jeden powiat – dąbrowski – z woj. małopolskiego. Pięć kolejnych powiatów należy do woj. mazowieckiego, przy czym są one znacznie oddalone od Warszawy. Są to powiaty: ostrołęcki i makowski (północna część województwa) oraz przysuski, szydlowiecki i zwolenński (południowa część województwa). Grupę dopełniają po dwie jednostki z województw podkarpackiego, podlaskiego i świętokrzyskiego.

Wśród 20 powiatów o najniższej stopie ubóstwa znajdują się wyłącznie miasta na prawach powiatu oraz jednostki bezpośrednio z nimi graniczące. Na pierwszym miejscu plasuje się Warszawa (6,9%), na drugim – Poznań (10,9%), a na trzecim – graniczący z Warszawą powiat pruszkowski (11,2%). Dalsze pozycje zajmują miasta na prawach powiatu: Gdynia i Sopot (po 11,7%) oraz Opole, Rzeszów i Olsztyn (po 12,8%).

Mapa terytorialnego zróżnicowania stopy i głębokości ubóstwa w przekroju powiatów w 2011 r.



Uwaga. W nawiasach okrągłych podano liczbę powiatów.
 Źródło: opracowanie własne na podstawie danych EU-SILC 2011 i NSP 2011.

Głębokość ubóstwa jest w dużym stopniu skorelowana ze stopą ubóstwa ($r = 0,99$). Najmniejsze ubóstwo wśród osób ubogich obserwuje się w Warszawie (2,6%), a następnie w powiecie pruskowskim i Sopocie (po 3,8%) oraz powiecie poznańskim i Poznaniu (po 3,9%). Największe wartości wskaźnika głębokości ubóstwa odnotowano w tych samych powiatach, w których występowała wysoka stopa ubóstwa.

Na podstawie wartości wskaźników ubóstwa oszacowanych na szczeblu powiatowym oraz macierzy sąsiedztwa obliczono także statystykę Morana I. Stwierdzono istnienie dodatniej korelacji przestrzennej – statystyka wyniosła odpowiednio 0,45 dla stopy ubóstwa i 0,47 dla głębokości ubóstwa. Istnienie umiarkowanie dodatniej autokorelacji przestrzennej świadczy o podobieństwie wartości sąsiadujących.

4.3. Zbieżność wyników estymacji pośredniej z danymi z rejestrów

W sytuacji gdy wartości wskaźników ubóstwa w populacji nie są znane, w ocenie ich poprawności można posłużyć się zmiennymi proxy (Zhang i Giusti, 2016). Wybór tych cech powinien być uzasadniony, a pomiędzy analizowaną cechą a zmienną proxy musi występować związek. Osoby, które korzystają ze świadczeń pomocy społecznej (zmienna proxy), są ubogie (zmienna analizowana). To samo można zauważyć w przypadku osób otrzymujących zasiłek dla bezrobotnych, których odsetek stanowi zmienną proxy dla liczby osób bezrobotnych. Trzeba jednak zauważyć, że nie zachodzi relacja odwrotna – nie wszystkie osoby bezrobotne otrzymują zasiłek (Fenton, 2013).

Oszacowania stopy i głębokości ubóstwa porównano z danymi na temat bezrobocia rejestrowanego oraz korzystania z pomocy społecznej. Zmienne te pochodziły z rejestrów, a zatem nie były obciążone błędem losowym, dzięki czemu można je uznać za precyzyjne miary porównawcze. Ponadto są to cechy ściśle związane ze zjawiskiem ubóstwa (GUS i US w Łodzi, 2017).

Oszacowania bezpośrednie są w bardzo małym stopniu skorelowane ze wskaźnikami z zakresu bezrobocia ($r = 0,198$) i pomocy społecznej ($r = 0,324$). Zastosowanie estymacji pośredniej bazującej na podejściu jednostkowym przynosi istotną poprawę. Porównanie tych oszacowań ze stopą bezrobocia długotrwałego pokazuje istotny wzrost wartości w porównaniu do estymacji bezpośredniej – w przypadku metody EB współczynnik korelacji liniowej Pearsona wynosi $r = 0,641$. Największe wartości współczynników obserwuje się dla cechy zasięgu korzystania ze środowiskowej pomocy społecznej ogółem – $r = 0,747$.

Wyniki estymacji świadczą o tym, że oszacowania pośrednie uzyskane przy zastosowaniu podejścia jednostkowego na poziomie powiatów są w dużo większym stopniu zbieżne z rzeczywistymi wartościami niż oszacowania bezpośrednie. Większe podobieństwo oszacowań i wskaźników opracowanych na podstawie rejestrów administracyjnych wskazuje na lepszą jakość wyników otrzymanych metodami statystyki małych obszarów.

5. Podsumowanie

Wartości wskaźników ubóstwa publikowane są dla całego kraju oraz w przekroju regionów i wybranych grup społeczno-demograficznych. Wielkość próby, jak również wykorzystywana obecnie metoda szacunku – estymacja bezpośrednia – z powodu dużych błędów oszacowań nie pozwalają na publikację wyników na niższych poziomach agregacji. Ponadto z roku na rok coraz więcej osób odmawia udziału w badaniach reprezentacyjnych. Wskaźnik kompletności w przypadku Europejskiego Badania Dochodów i Warunków Życia (EU-SILC) w 2006 r. wyniósł 69,9%, a w 2016 r. – już tylko 48,4% (GUS, 2008, 2017). Niemniej równocześnie z obserwowanym spadkiem wartości wskaźników realizacji badań reprezentacyjnych rosną potrzeby odbiorców informacji. Oczekuje się danych dla szczegółowych przekrojów oraz coraz mniejszych jednostek administracyjnych czy terytorialnych.

W badaniu omawianym w artykule do oszacowania wskaźników ubóstwa na poziomie powiatów zastosowano estymację pośrednią – metodę empiryczną bayesowską (EB), która polega na generowaniu pseudopopulacji metodą Monte Carlo. W tym celu wykorzystano dane z EU-SILC oraz Narodowego Spisu Powszechnego Ludności i Mieszkań (NSP). Metoda EB bazuje na danych jednostkowych, które lepiej opisują ubóstwo gospodarstw domowych niż dane zagregowane. Tym samym parametry liniowego modelu mieszanego szacowane są na podstawie większej liczby obserwacji. Przyjęcie za zmienną zależną dochodu ekwiwalentnego gospodarstwa umożliwia szacowanie, oprócz wskaźników ubóstwa, dowolnego parametru opartego na dochodzie, np. współczynnika Giniego.

Zastosowanie metod statystyki małych obszarów umożliwiło uzyskanie precyzyjnych oszacowań stopy i głębokości ubóstwa na poziomie powiatów oraz oszacowanie wartości wskaźników ubóstwa także dla tych 16 jednostek terytorialnych, w których nie było żadnego reprezentanta populacji ubogich lub które nie znalazły się w próbie EU-SILC. Otrzymane wyniki istotnie zwiększają szczegółowość dostępnych informacji, zwłaszcza w przypadku wskaźnika głębokości ubóstwa, który publikowany jest wyłącznie w wybranych przekrojach społeczno-demograficznych. W celu merytorycznej oceny oszacowań zaproponowano wykorzystanie danych na temat bezrobocia i pomocy społecznej pochodzących z rejestrów administracyjnych. Wykazano, że korelacja pomiędzy tymi zmiennymi a oszacowaniami pośrednimi jest wyższa niż w przypadku oszacowań bezpośrednich.

Spis ludności stanowi podstawowe źródło danych o stanie i strukturze populacji kraju. Ze względu na koszty i zakres badanie to jest realizowane co dekadę. Wyniki omawianego badania, w którym wykorzystano dane z NSP 2011, nie odnoszą się zatem do sytuacji gospodarstw domowych w Polsce aktualnej w czasie publikacji artykułu. Od 2011 r. wdrożono wiele programów pomocy społecznej skierowanych

do gospodarstw domowych, takich jak „Rodzina 500+”, „Aktywne formy przeciwdziałania wykluczeniu społecznemu”, „Program Operacyjny Pomoc Żywnościowa”, „Karta Dużej Rodziny” oraz wiele innych. Mają one na celu m.in. ograniczenie ubóstwa. Jędrzejczak i Pekasiewicz (2020), na podstawie badania budżetów gospodarstw domowych, wskazują na istotną poprawę sytuacji dochodowej rodzin wielodzietnych w latach 2015 i 2016. Ocenę wpływu transferów społecznych na problem ubóstwa można znaleźć m.in. w pracach Brzezińskiego i Najsztuba (2017), Szarfenberga (2018) i Szulca (2019).

Zastosowanie zaproponowanej metodyki oraz wykorzystanie danych ze spisu powszechnego, który będzie prowadzony w 2021 r., powinno umożliwić porównanie poziomu ubóstwa w gospodarstwach domowych w latach 2011 i 2021. Oszacowanie wskaźników ubóstwa z wykorzystaniem metody EB dla pozostałych lat wymaga dostępu do danych jednostkowych, np. z rejestrów administracyjnych. Dane z rejestrów rządowych mogą być wykorzystane także w celu korekty rozkładu dochodów, ponieważ osoby osiągające bardzo wysokie dochody zwykle odmawiają udziału w badaniach ankietowych (Bukowski i Novokmet, 2019).

Bibliografia

- Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. DOI: 10.18637/jss.v067.i01.
- Battese, G. E., Harter, R. M., Fuller, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83(401), 28–36. DOI: 10.1080/01621459.1988.10478561.
- Bedi, T., Coudouel, A., Simler, K. (2007). *More Than a Pretty Picture: Using Poverty Maps to Design Better Policies and Interventions*. Washington: The World Bank.
- Biecek, P. (2011). *Analiza danych z programem R: Modele liniowe z efektami stałymi, losowymi i mieszanymi*. Warszawa: Wydawnictwo Naukowe PWN.
- Brzeziński, M., Najsztub, M. (2017). The impact of „Family 500+” programme on household incomes, poverty and inequality. *Polityka Społeczna*, (1), 16–25.
- Bukowski, P., Novokmet, F. (2019). Between Communism and Capitalism: Long-Term Inequality in Poland, 1892–2015. *CEP Discussion Paper*, (1628).
- Chambers, R., Tzavidis, N. (2006). M-quantile models for small area estimation. *Biometrika*, 93(2), 255–268. DOI: 10.1093/biomet/93.2.255.
- Chandra, H., Chambers, R., Salvati, N. (2012). Small area estimation of proportions in business surveys. *Journal of Statistical Computation and Simulation*, 82(6), 783–795. DOI: 10.1080/00949655.2011.554834.
- Dehnel, G. (2010). *Rozwój mikroprzedsiębiorczości w Polsce w świetle estymacji dla małych domen*. Poznań: Wydawnictwo Uniwersytetu Ekonomicznego.
- EEC. (1985). Council Decision of 19 December 1984 on specific community action to combat poverty. *Official Journal of the European Communities*, 28(L 2), 24–25.

- Fenton, A. (2013). *Small-area measures of income poverty*. London: Centre for Analysis of Social Exclusion, London School of Economics.
- Foster, J., Greer, J., Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica*, 52(3), 761–776. DOI: 10.2307/1913475.
- Gołata, E. (2004). *Estymacja pośrednia bezrobocia na lokalnym rynku pracy*. Poznań: Wydawnictwo Akademii Ekonomicznej.
- González-Manteiga, W., Lombardía, M., Molina, I., Morales, D., Santamaría, L. (2008). Analytic and bootstrap approximations of prediction errors under a multivariate Fay-Herriot model. *Computational Statistics and Data Analysis*, 52(12), 5242–5252. DOI: 10.1016/j.csda.2008.04.031.
- Graf, M., Marín, J. M., Molina, I. (2019). A generalized mixed model for skewed distributions applied to small area estimation. *Test*, 28(2), 565–597. DOI: 10.1007/s11749-018-0594-2.
- Guadarrama, M., Molina, I., Rao, J. N. K. (2016). A comparison of small area estimation methods for poverty mapping. *Statistics in Transition new series and Survey Methodology*, 17(1), 41–66. DOI: 10.21307/stattrans-2016-005.
- GUS. (2008). *Dochody i warunki życia ludności Polski (raport z badania EU-SILC 2006)*. Warszawa: Główny Urząd Statystyczny.
- GUS. (2017). *Dochody i warunki życia ludności Polski (raport z badania EU-SILC 2015)*. Warszawa: Główny Urząd Statystyczny.
- GUS. (2019). *Dochody i warunki życia ludności Polski – raport z badania EU-SILC 2018*. Warszawa: Główny Urząd Statystyczny.
- GUS. (2020). *Sytuacja gospodarstw domowych w 2019 r. w świetle wyników badania budżetów gospodarstw domowych*. Informacja sygnałowa z dnia 29.05.2020. Pobrane z: https://stat.gov.pl/files/gfx/portalinformacyjny/pl/defaultaktualnosci/5486/3/19/1/sytuacja_gospodarstw_domowych_w_2019_r_w_swietle_wynikow_badiani_budzetow_gospodarstw_domowych.pdf.
- GUS, US w Łodzi. (2013a). *Jakość życia, kapitał społeczny, ubóstwo i wykluczenie społeczne w Polsce*. Warszawa: Główny Urząd Statystyczny.
- GUS, US w Łodzi. (2013b). *Ubóstwo w Polsce w świetle badań GUS*. Warszawa: Główny Urząd Statystyczny.
- GUS, US w Łodzi. (2017). *Jakość życia w Polsce w 2015 roku: Wyniki badania spójności społecznej*. Warszawa: Główny Urząd Statystyczny.
- Haughton, J., Khandker, S. R. (2009). *Handbook on Poverty and Inequality*. Washington: The World Bank.
- Horvitz, D. G., Thompson, D. J. (1952). A Generalization of Sampling Without Replacement From a Finite Universe. *Journal of the American Statistical Association*, 47(260), 663–685. DOI: 10.2307/2280784.
- Jędrzejczak, A., Pekasiewicz, D. (2020). Changes in Income Distribution for Different Family Types in Poland. *International Advances in Economic Research*, 26(2), 135–146. DOI: 10.1007/s11294-020-09785-1.
- Kurowski, P. (2002). *Kosztyki minimum socjalnego i minimum egzystencji – dotychczasowe podejście*. Pobrane z: https://www.ipiss.com.pl/wp-content/uploads/downloads/2012/08/rola_funk_min_soc_egz.pdf.
- Molina, I., Marhuenda, Y. (2015). Sae: An R Package for Small Area Estimation. *The R Journal*, 7(1), 81–98. DOI: 10.32614/rj-2015-007.

- Molina, I., Rao, J. N. K. (2010). Small area estimation of poverty indicators. *The Canadian Journal of Statistics*, 38(3), 369–385. DOI: 10.1002/cjs.10051.
- Panek, T. (2010). Multidimensional approach to poverty measurement: Fuzzy measures of the incidence and the depth of poverty. *Statistics in Transition new series*, 11(2), 361–380.
- Panek, T. (2011). *Ubóstwo, wykluczenie społeczne i nierówności: Teoria i praktyka pomiaru*. Warszawa: Oficyna Wydawnicza Szkoła Główna Handlowa.
- Rao, J. N. K., Molina, I. (2015). *Small Area Estimation*. Hoboken, New Jersey: John Wiley & Sons.
- Rojas-Perilla, N., Pannier, S., Schmid, T., Tzavidis, N. (2020). Data-driven transformations in small area estimation. *Journal of the Royal Statistical Society: Series A: Statistics in Society*, 183(1), 121–148. DOI: 10.1111/rssa.12488.
- Szarfenberg, R. (2018). Ubóstwo i pogłębiona deprywacja materialna rodzin w kontekście wdrożenia programu „Rodzina 500 plus”. *Polityka Społeczna*, 45(1), 11–17.
- Szulc, A. (2019). Transfery socjalne w Polsce w kontekście ubóstwa monetarnego i wielowymiarowego. *Wiadomości Statystyczne. The Polish Statistician*, (3), 7–26. DOI: 10.5604/01.3001.0013.8519.
- Szymkowiak, M., Młodak, A., Wawrowski, Ł. (2017). Mapping poverty at the level of subregions in Poland using indirect estimation. *Statistics in Transition new series*, 18(4), 609–635. DOI: 10.21307/stattrans-2017-003.
- Szymkowiak, M. (red). (2015). *Pomiar ubóstwa na poziomie powiatów (LAU 1) – etap II*. Jachranka: Centrum Badań i Edukacji Statystycznej GUS.
- Ulman, P., Ćwiek, M. (2020). Measuring housing poverty in Poland: a multidimensional analysis. *Housing Studies*, 1–19. DOI: 10.1080/02673037.2020.1759515.
- Van der Weide, R. (2014). *A Review of Empirical Bayes Prediction Using Sampling Weights with Poverty Mapping in Mind*. Washington: The World Bank.
- Wawrowski, Ł. (2016). The Spatial Fay-Herriot Model in Poverty Estimation. *Folia Oeconomica Stetinensia*, 16(2), 191–202. DOI: 10.1515/fofi-2016-0034.
- Wilak, K. M. (2014). Strukturalne modele szeregów czasowych w estymacji stopy bezrobocia w dezagregacji na województwa, płeć i wiek. *Przegląd Statystyczny*, (4), 409–431.
- Wolter, K. (2007). *Introduction to variance estimation*. New York: Springer.
- Zhang, L.-C., Giusti, C. (2016). Small Area Methods and Administrative Data Integration. W: M. Pratesi (red.), *Analysis of Poverty Data by Small Area Estimation* (s. 61–82). Hoboken, New Jersey: John Wiley & Sons.
- Żądło, T. (2015). *Statystyka małych obszarów w badaniach ekonomicznych: Podejście modelowe i mieszane*. Katowice: Wydawnictwo Uniwersytetu Ekonomicznego.